Albert-Lőrincz Csanád

Módszertani útmutató kvalitatív és kvantitatív kutatásokhoz

Presa Universitară Clujeană

Csanád Albert-Lőrincz

•

MÓDSZERTANI ÚTMUTATÓ KVALITATÍV ÉS KVANTITATÍV KUTATÁSOKHOZ

Kézikönyv társadalomtudományi szakos hallgatóknak

Támogatók:



Megvalósult
a Magyar Kormány
támogatásával



Miniszterelnökség
Nemzetpolitikai Államtitkárság



Bethlen Gábor
Alap





Csanád Albert-Lőrincz

MÓDSZERTANI ÚTMUTATÓ KVALITATÍV ÉS KVANTITATÍV KUTATÁSOKHOZ

Kézikönyv társadalomtudományi szakos hallgatóknak

PRESA UNIVERSITARĂ CLUJEANĂ

2023

Referenți științifici:

Prof. univ. dr. Georgina Szilágyi

Prof. univ. dr. Gavril Flora

ISBN 978-606-37-2017-8

Universitatea Babeş-Bolyai
Presa Universitară Clujeană
Director: Codruţa Săcelean
Str. Hasdeu nr. 51
400371 Cluj-Napoca, România
Tel./fax: (+40)-264-597.401
E-mail: editura@editura.ubbcluj.ro
http://www.editura.ubbcluj.ro/

TARTALOMJEGYZÉK

SZERZŐI ELŐSZÓ

A szakdolgozatra készülő diákok gyakran szembesülnek a témaválasztás kapcsán bizonytalansággal, majd a kutatás előkészítésének és megtervezésének nehézségeivel. Több helyről is kaphatnak segítséget: fellapozhatják az előző tanévekben sikeresen megvédett dolgozatok könyvtárakban vagy egyetemi tanszéki szobákban fellelhető példányait, útmutatást kérhetnek tanáraiktól, vagy ismételten kézbe vehetnek néhány olyan módszertani könyvet, amelyekkel tanulmányi éveik során ismerkedhettek meg.

Jelen kézikönyv ez utóbbi lehetőséget szolgálja, útmutató jelleggel összegzi a kutatás előkészítésének és kivitelezésének fontosabb mozzanatait, mint például a témaválasztást, tervezést, adatgyűjtést és feldaolgozást, valamint a védés menetét.

Saját kutatási tapasztalataimmal, az általam vezetett és/vagy elbírált szakdolgozatok tanulságait összefoglalva szeretném segíteni a hallgatók államvizsgára való felkészülését. Tizenöt évnyi oktatói és kutatói tapasztalatomat osztom meg a társadalomtudományok területén kutatni vágyó egyetemi hallgatókkal; a szakdolgozattal kapcsolatos munkájukat szándékszom megkönnyíteni. Mivel a kutatás módszertana általánosan alkalmazható a humántudományok minden ágazatában, tehát egyetlen szakiránynak sem sajátja, ezért jelen kézikönyv minden olyan hallgatónak hasznára válhat, aki az emberi viselkedést, a társadalom működését, vagy akár a szervezeti kultúra bizonyos aspektusait szeretné kutatni.

Célom, hogy az egyetemi hallgatók által végzett kutatásokhoz módszertani ismereteket nyújtsak, olyan logikai struktúrába foglalva, amely a kutatások megvalósítását hatékonyabbá teheti, valamint a ráfordított időt minimalizálja. Továbbá, a kutatás és a kutatási eredmények bemutatása során gyakran felszínre kerülő hibákra is szeretném felhívni a az egyetemi ciklust lezáró dolgozatok szerzőinek figyelmét. A hangsúlyt a „mit kutassunk?", „miért?" és „hogyan?" gyakorlati megvalósítására helyezem, beleértve az online források felhasználására vonatkozó útmutatásokat is.

Arra törekedtem, hogy *Módszertani útmutató a kvantitatív és a kvalitatív kutatásokhoz* című kézikönyvem könnyen áttekinthető rendszerezés, hasznos útmutató legyen, következésképpen nem helyettesíti és nem is fedi le teljes mértékben az alapképzés tanrendjében szereplő összes módszertani temati-

kát. Azonban felhívnám a hallgató figyelmét, hogy mindig a témavezető tanár útmutatásait tekintse elsődlegesnek a kutatás során, mivel minden oktatónak és kutatónak sajátos szakami elvárásai lehetnek és a különböző intézmények követelményrendszere sem azonos a szakdolgozatokat illetően.

A szerző

1. MI IS AZ A KUTATÁSMÓDSZERTAN?

A világ megismerése és megértése az ember egyik alapvető igénye. A természet fürkészése, az összefüggések feltárására vonatkozó megfigyelések és kísérletek már az ókortól ismertek. Az emberiség fejlődése során annál messzibbre jutott, minél több és minél alaposabb ismeretekre tett szert, minél pontosabb és kifinomultabb eredményekhez, összefüggésekhez, felfedezésekhez jutott. A természet és a társadalom fokozatos megismerése egy olyan folyamat, melyet bizonyos módszertani keretek kísérnek. Időben nemcsak a jelenségekről és a folyamatokról való tudásunk gyarapodott, hanem a megismerés mikéntje, eszközei is fokozatosan bővültek: ez a folyamat a kezdeti megfigyelésen alapuló módszerektől a matematikai-statisztikai módszerek alkalmazása felé bővült. Tehát a társadalomtudományok fejlődésének velejárója, hogy a módszerek is sokasodtak és diverzifikálódtak, sőt a vizsgált területek is folyamatos bővültek az egyre bonyolultabb összefüggések felé. A társadalomtudományok fejlődését a többi tudományág fejlődése is nagyban befolyásolja. A XX. század áttörést hozott az emberi viselkedés megismerésében, a matematikában, fizikában, orvostudományban, mivel az addigi ismereteink jelentősen kibővültek. Mindez új irányzatok és tudományágak létrejöttét eredményezte. A világunk megismerésének folyamata jelen korunkban felgyorsult és egy újabb jelentős fejlősédi szakaszba lépett, melyet a különböző matematikai algoritmusoknak és a mesterséges intelligencia rohamos fejlődésének köszönhetünk.

A kutatás módszertana, mint ahogyan a neve is sugallja, olyan módszerek és eszközök összessége, amelyek lehetővé teszik egy társadalmi probléma vagy jelenség vizsgálatát, elemzését. Ezen módszerek és eszközök általánosan alkalmazhatóak a humántudományokban, tehát nem a szociális munka, szociológia, pszichológia, gazdaságtan vagy orvostudomány sajátos eszközei.

1.1. A kutatás etikai kihívásai az elméletben

A kutatást egy öngerjesztő folyamatként is értelmezhetjük, mivel minél többet tudunk, annál többet és több irányba kutatunk. A pontos eredményekhez szükségünk van egy olyan keretre, amely tartalmazza az elérhető eszközöket és módszereket, valamint ezek alkalmazásának szabályait, korlátait. Bár a tudományos kutatás szabadsága mára egy széles körben elfogadott alapjog, bizonyos etikai és jogi vonatkozásokat tiszteletben kell tartanunk.

Az erkölcsi korlátok és fenntartások nélkül, pusztán a tudomány nevében végzett kutatások eredményzeték például a hírhedt náci emberkísérleteket. A Második Világháború utáni nürnbergi orvosper egyértelművé tette, hogy az emberkísérleteket szigorúan korlátozni kell. Az ítélethirdetés eredményeként született meg 1947-ben a Nürnbergi Kódex, mely 10 alapelvet vezettet be az orvosi etikában. Nemzetközi színtéren az emberi jogok térnyerése további etikai szabályokat vont maga után; elvi irányokat fektet le például az Orvos Világszövetség (WMA) Helsinki Nyilatkozata 1964-ből az embereken végzett orvosi kutatások etikai alapelveiről.

A tudomány „érdeke" tehát nem lehet fontosabb, mint az emberi jogok védelme, viszont ebből a megállapításból napjainkra egy újabb korlát nőtte ki magát, amivel minden kutató szembesülni fog: ez a korlát az adatvédelmi megfelelőség. Az adatvédelem az EU-ban a magánélethez való jogból származtatott olyan érték, mely alapjogi szintre lett emelve, így a kutatás etikájának szerves részévé vált.

A társadalmi szférát jelenleg meghatározó legfőbb érték az emberi jogok védelme, melynek létrejötte a Második Világháború alatt elkövetett pusztítás következménye. Ez az értékrend egyre kiterjedtebb és részletgazdagabb lett:

– elsőként az ENSZ 1948-ban elfogadta az Emberi Jogok Egyetemes Nyilatkozatát;

– az Európa Tanácsa 1950-ben elfogadta az Európai Emberjogi Egyezményt és a következő években ennek 16 Kiegészítő jegyzőkönyvét;

– további emberjogi ENSZ dokumentumok az 1966-os Polgári és Politikai Jogok Nemzetközi Egyezségokmánya;

– a szintén 1966-os Gazdasági, Szociális és Kulturális Jogok Nemzetközi Egyezségokmánya;

– az 1989-es Gyermekek Jogairól szóló Egyezmény;

– az EU szintjén pedig a 2000-ben kihirdetett Európai Unió Alapjogi Chartája;

– vagy Az Európai Parlament és a Tanács 2016/679 számú Általános Adatvédelmi Rendelete.

Bármely kutatásban, területtől függetlenül, a fenti dokumentumokban kikiáltott alapjogokat tiszteletben kell tartani, sőt ezen túlmenően további engedélyekre is szükségünk lehet.

Az alanyok kutatásban való részvételhez való hozzájárulása is szükséges minden esetben, de ez a beleegyezés lehet akár hallgatólagos is, amikor például az alany kitölt egy kérdőívet. A kiskorúak vagy cselekvőképtelen felnőttek esetén szükséges a szülői engedély vagy a gyám beleegyezése, ami körülményessége miatt gyakran ellehetetleníti a velük végzett adatfelvételt. Ez a nézőpont elsősorban a hazai és nemzetkzöi kutatói közösségek gyakorlatából fakad, például „a Magyar Szociológiai Társaság Etikai Kódexe (2001) nem tér ki a gyerekekkel történő kutatásra" (Kutrovtz, 2016, 95), mely átveszi a Nemzetközi Szociológiai Társaság (ISA) Etikai Kódexét. Továbbá, amennyiben egy adott intézmény vagy

szervezet keretében végzünk adatgyűjtést, akkor az intézmény vezetőjének előzetes engedélyét kell beszereznünk.

Összegzésként elmondható, hogy a társadalomtudományokban rendszerint a szakmai szervezetek, társaságok és kamarák teszik közzé az adott szakma etikai kódexét, illetve a kutatás etikáját, de az általánosan elvárt kutatási elvek a szakirodalomban is körvonalazódhatnak.

1.2. A kutatás etikai kihívásai a gyakorlatban

A kutatás komplex jellegéből adódóan a gyakorlatban is felmerül egy sor etikai kérdés és dilemma. Ezek feloldása sokszor nem egyszerű, de az alapelvek mentén bizonyossággal találunk megoldást. A felmerülő kérdések az alábbi témakörök által közelíthetőek meg:

– tiszteletben tartottuk-e az emberi jogokat?

– betartottuk-e az etikai elveket?

– alkalmaztuk-e a „nil nocere" vagyis „ne árts"[1] elvet?

Az 1.1.-es alfejezetben említett kutatásetikai szempontok a következő elvekkel bővülnek ki:

a. A kutatás valós adatokon alapuljon;

b. Az adatgyűjtés a módszertanban leírtak szerint történjen;

c. A szükséges engedélyek beszerzése;

d. Az alanyok előzetes tájékoztatása és a részvételi hozzájárulásuk megszerzése;

e. Az alanyok identitásának védelme;

f. A kutatás színterének helyet adó intézmény érintettsége.

Ezeknek a szempontoknam az érvényesülését a kutatás konkrét terve és menete határozza meg. Ezen alapelvek és szempontok némelyike összefügg, egyikből következik a másik, ezért tekintsük őket egységükben.

a. A kutatás valós adatokon alapuljon

Alapvető elvárás, hogy az eredmények valós személyekkel folytatott interjúkon, ténylegesen lekérdezett kérdőíveken, valós esetszámokon, egyéb, valóban alkalmazott eszközökön alapuljon. Ezen alapelv megszegése a hallgató kizárásával járhat adott vizsgaszesszióból. Amennyiben az adatok valódisága felől kételyek merülnénk fel, a témavezető tanár, vagy a bizottság bármely tajgja bekérheti a lealkalmazott kérdőíveket, helyszínen készített interjúvázlatokat vagy egyéb, az adatgyűjtéssel kapcsolatos jegyzeteket, fájlokat. Bár a kutatást a hallgató végzi, ennek felügyelete a vezető tanár felelőssége is.

1 A „nil nocere" alapelv az orvosi filozófiából származtatható, de a többi társadalomtudományban is alkalmazzuk. Például a szociális munka szakmai deontológiája is elismeri.

Az adathamisítás vagy szándékos elferdítés bármilyen formája etikai vétség és tilos, ráadásul szükségtelen, mivel az is eredménynek számít, ha nem találtunk két jelenség között összefüggést, nem találtuk meg az ok-okozati kapcsolatot, esetleg arra a következtetésre jutottunk, hogy a témánkon belül további vizsgálatokra van szükség.

Jelen alapelv például olyankor sérülhet, amikor egy kutatást megrendelésre készítenek, valamint a kutatást finanszírozó szervezet elvár egy bizonyos eredményt. Megemlíthetnénk a 60-as években készült, a dohányzás káros hatásait vizsgáló kutatásokat, ahol a megrendelő dohányipari cégek (és az érintett kutatók is) több évig rejtegették a valós kockázatokat, amiatt, hogy ha fény derült volna a valós adatokra, akkor nyereségtől estek volna el. A valós kockázati hatásokat bár későbbi tanulmányok napvilágra hozták, a szándékos félrevezetés és adathamisítás okozta egészségügyi és társadalmi károk mértéke kétségkívül jelentősnek bizonyult.

b. Az adatgyűjtés a módszertanban leírtak szerint történjen

A kutatás egy előre elkészített terv szerint valósul meg, amiben már az adatgyűjtésre vonatkozóan is fel kell vázolnunk az elképzelésünket, meg kell jelölnünk a vizsgálati alanyainkat (kvalitatív módszer esetén) vagy mintánkat (kvantitatív módszer esetén). Ha eltérünk a tervtől és rögtönzünk, az eredmények megbízhatatlanná válhatnak. Például egy kérdőíves kutatás során, ha a minta egy adott város összes iskolájával, ezek különböző szakirányaival és osztályaival számol, akkor precíz adatokat kaphatunk az adott település oktatási helyzetéről. Ha azonban az adatgyűjtés során nincs mód az összes eset vizsgálatára, mintavétellel juthatunk megbízható eredményekhez. Ha az adatfelvétel nem a tervezett reprezentatív mintavétel[2] szerint történik, az eredmények szinte biztosan torzítani fognak és a megbízhatóságuk kérdésessé válik. Amennyiben ugyanazt a kutatást megsmételnénk, a kapott eredmények nem lennének reprodukálhatóak, ezáltal a kutatás tudományos értéke megkérdőjeleződik.

A gyakorlatban sajnos előfordulnak olyan esetek, amikor nem sikerül az előre eltervezett minta szerint bevonnunk az alanyokat. Például nem kapunk engedélyt, hogy a mintában szereplő intézményben/iskolában/osztályban felvegyük az adatokat. Ezért javasolt, hogy mielőtt elkezdenénk a terepmunkát, a részleges adatfelvételt, előre egyeztessük az illetékes személyekkel a megvalósíthatóságot. Amennyiben akadályba ütköznénk, még idejében újratervezhető a minta.

A megfelelő körültekintés ellenére is adódhatnak előre nem látható helyzetek, ezért szükség esetén a mintánkat pótolhatjuk a tervezettel azonos jellem-

2 A reprezentatív mintavétel röviden azt jelenti, hogy a kutatásban résztvevő személyek, alanyok bizonyos szempontok alapján leképezik a vizsgált populáció (alapsokaság) egészét, annak bizonyos jellemzőit. Bővebb leírást a 11.2.1 alfejezet tartalmaz.

zőkkel bíró alanyokkal, hogy arányaiban ne sérüljön a populáció minta általi leképezése, reprezentativitása. Viszont ennek a tényét mindeképpen jeleznünk kell a kutatásban.

c. A szükséges engedélyek beszerzése

Amennyiben egy intézményben végezzük a kutatást, adatfelvételt, előzetes írásos engedélyre lehet szükségünk, amelyet az intézmény vezetőjénél vagy igazgatójánál kérvényezhetünk. A kérvényhez csatolni szokták az interjúválzatot, illetve kérdőívet. Az interjúvázlatban a főbb témaköröket rögzítjük, azokat a kérdéseket fogalmazzuk meg, amelyekre választ szeretnénk kapni.

Az engedély beszerzése a gyakorlatban akkor lehet nehezebb, ha nincs előzetes kapcsolatunk az intézménnyel. Javasolt inkább olyan intézmények kiválasztása, ahol már ismernek, például önkéntesként ott dolgoztunk, szakmai gyakorlatot végeztünk vagy alkalmazásban állunk/álltunk. Így a bizalmi viszony eleve adott és valószínűbb, hogy az adatfelvétel során együttműködőbbek és őszintébbek lesznek az alanyok.

Nem ritkák azok az esetek sem, amikor a kutatónak nem engedélyezik egy adott intézményben az adatgyűjtést. Ilyenkor az intézmény vezetője közli a kutatással kapcsolatos kifogásait vagy hárít, arra hivatkozva, hogy neki is „felsőbb" engedélyeket kell kérnie. Ennek oka, hogy az intézmény vezetője nem szeretné sem magát, sem az általa vezetett intézményt összefüggésbe hozni a kutatás által esetlegesen felszínre hozott problémákkal. Továbbá a nem vezető beosztású munkatársaktól is nehéz valós információkhoz jutni, mivel ők is tarthatnak a feletteseiktől, ezért ők sem szeretnének az intézményen belüli problémákról beszélni.

A kutatás intézményi jóváhagyása még nem feltétlenül jelenti azt, hogy az intézményben dolgozó munkatársak őszinték lesznek, ezért arra kell számítanunk, hogy a valós helyzetnél jobb színben fogják feltüntetni a helyzetet.

Az adatgyűjtés gyakorlati nehézségei kapcsán néhány példát említenék: Egy kutató, bár rendelkezett előzetes tankerületi engedéllyel, mégsem jutott be egy teológia szakirányú középiskolába, mivel az igazgató úgy vélte, hogy „rossz" a kérdőív. Azt kifogásolta, hogy a kérdőív tartalmaz egy olyan kérdést, hogy a tanulók követtek-e el, illetve szenvedtek-e el zaklatást társaik részéről. Sajnálatos módon a zaklatás egy létező jelenség, viszont az igazgató nem szeretett volna a kutatás miatt a szülők előtt esetlegesen magyarázkodni. Hasonló eset történt egy másik kutatás kapcsán is, ahol a kérdőíves lekérdezést az igazgató engedélyezte, viszont egy osztályfőnök ennek ellenére nem engedte az adatfelvételt, szintén a szülők felháborodásától tartva. Mindkét esetben anonim kérdőívekről volt szó, tehát nem állt fenn a tanulók személyiségi jogsérelmének kockázata.

Harmadik példaként egy orvosokkal végzett doktori kutatást említenék, mely a gyermekjogok érvényesülését vizsgálta a korházi ellátásban. Bár a kutató rendelkezett a szükséges engedélyekkel, az orvosok nem nyíltak meg, és kivétel nélkül azt állították, hogy az intézményben minden gyermekjogot betartanak. Mivel az adott kutatásba szülőket és gyermekeket is bevontak, az orvosok optimista válaszaihoz képest sokkal árnyaltabb helyzetet sikerült feltárni.

Bármilyen kutatásról legyen szó, tudatában kell lennünk, hogy az intézményi kultúra hatással van a válaszokra, így az eredményekre is: minden egyén hajlamos valamennyire azonosulni a munkahelyével, az intézményen belüli feladat- vagy szerepköreivel, hivatásával, így annak körülményeit szubjektíven jobbnak látja, mint amit a valóság mutat.

Ha egy intézmény munkatársa nem válaszol meg egy kérdést, hárít, vagy nyilvánvalóan nem tárja fel (eltitkolja) a problémákat, akkor más módon, más eszközökkel kell az adatokat megszereznünk.

d. Az alanyok előzetes tájékoztatása és részvételi hozzájárulásuk megszerzése

A személyekkel folyatott kutatások alkalmával rendszerint a válaszadókat előzetesen tájékoztatni kell a kutatás témájáról és célkitűzéséről, valamint részvételi hozzájárulásukat kell kérni, még anoním adatfelvétel esetén is. Ezt megtehetjük az interjúk megkezdése előtt a bemutatkozás részeként, valamint kérdőíves felmérések esetén a fenti tájékoztatást feltüntethetjük a kérdőív elején. Ha személyesen vesszük fel a kérdőíves válaszokat, elegendőnek tartjuk a szóbeli tájékoztatást.

Kivételes esetben, ha a kutatás tényének közlése befolyásolná az alanyok viselkedését, eltekinthetünk a tájékoztatástól. Ilyen lehet például egy csoport megfigyelése, amikor bármilyen figyelemfelkeltő beavatkozás módosíthatja az eredményeket. Két dologra azonban ilyenkor is figyelnünk kell, az első az alanyok valós személyazonosságának elfedése, vagyis az anonimitás biztosítása, a második az előzetes intézményi engedélyek beszerzése. További példa, amikor eltekinthetünk az előzetes tájékoztatástól, az A/B típusú tesztek alkalmazása. Ennek során az alanyok két lehetőség közül a számukra jobban tetszőt jelölik meg (lehet például két hasonló termék, két szín vagy akár két különböző design). Véleményünk szerint olyankor sem szükséges az alanyok tájékoztatása, amikor a szociális médián fellelhető publikus információkat, profilokat elemezzük. Amennyiben Big Data-n alapuló kutatást végzünk, mint például a keresési kulcsszóelemzést, értelemszerűen magunk sem ismerhetjük az alanyok kilétét, mégis pontos, megbízható eredményeket kaphatunk.

e. Az alanyok identitásának védelme

Az identitás, vagy személyiség elfedése azt jelenti, hogy az alany anoním marad a kutatás során. Eleve nem rögzítünk személyes adatokat, mint például név, lakcím, telefonszám, e-mail cím, fotók stb. Ez a felsorolás nem kizárólagos, az alany minden olyan jellemzőjére kiterjed, amely azonosíthatóvá tehet egy adott személyt. Azonban a statisztikai feldolgozás érdekében szükségünk lehet bizonyos demográfiai, szocio-kulturális és szocio-gazdasági jellemzőkre, értelemszerűen ezek nem számítanak személyes adatnak. Mindig a konkrét kutatás keretei között kell mérlegelnünk, hogy milyen adatokra lesz feltétlenül szükségünk; a túlzott adatgyűjtést kerülnünk kell. A helyzetet bonyolítja az a tény, hogy önmagában egy kérdőív még nem jelent garanciát a teljes anonimitásra. Például, ha egy iskolai osztályban lealkalmazott kérdőívben rákérdezünk a szülők foglalkozására, feltehetően ki lehet következtetni néhány tanuló azonosságát. Ha az életkort is megkérdeztük, elképzelhető, hogy a legidősebb és/vagy legfiatalabb tanuló is azonosítható, amennyiben születési éve eltér a többiekétől.

Az anonimitást ezért nem a kérdőív, hanem a kutatás vonatkozásában kell garantálnunk. Ha az adatokat statisztikai elemzésnek vetjük alá, eleve megoldódik a fenti dilemma, mert az eredményeink általános jellemzőket és tendenciákat fognak tükrözni, melyből nem fejthető vissza a résztvevők személyazonossága.

Az anonimitás garantálásának további buktatója lehet, ha egy kérdőívet egy általunk ismert alanyhoz juttatjuk el vagy veszünk át tőle (online vagy személyesen). Konkrét példaként említjük, amikor egy diák egyetemi oktatókat keresett fel egy anonimnak szánt kérdőívével, melyben az oktatás során felmerülő problémákat szerette volna feltárni. A felkérést és a mellékelt kérdőívet az oktatók e-mail címére küldte ki, és ugyanígy várta volna a válaszokat is. Azonban az e-mail cím személyhez rendelhető, így az anonimitást nem tekintettük biztosítottnak és a kutatásban való részvételt visszautasítottuk. Tartalmilag is akadtak gondok a kérdőív kapcsán: bár a névre nem kérdezett rá, az akadémiai titulus és didaktikai fokozat, valamint a kar és tanszék valószínűsíthetővé tette volna a válaszadó személyét. Ezen túlmenően olyan kérdéseket is tartalmazott, amelyek az egyetemen tapasztalt oktatással kapcsolatos problémákat feszegették. A helyzet alapjában véve ugyanaz, mint a jelen fejezet c) pontjánál ismertetett példában, az orvosokkal végzett felmérés kapcsán, vagyis a problémákról feltehetően az oktatók nem fognak objektíven beszámolni.

Az oktatásügy – bármely más területhez hasonlóan – kutatható, csak a megfelelő eszközöket kell használni, illetve a megfelelő alanyokat kell felkeresni. Például igényelhetünk adatokat az intézménytől, felkereshetünk jelenlegi és volt diákokat, végezhetünk megfigyelést az óralátogatás és diáklétszám kapcsán stb.

f. A kutatás színterének helyet adó intézmény érintettsége

A legtöbbször a kutatást nem az utcán, köztereken végezzük, hanem egy adott intézményen belül. Az intézmények némelyikének van a kutatásra vonatkozó belső szabályzata, ezt nem árt ellenőrizni. Általában a kutatóktól előzetesen elkérik az interjúvázlatot, a kérdőív tervezetet, és megnézik, hogy az intézmény szempontjaiból rendben van-e? Ezért célszerű, hogy az intézményre vonatkozóan ne tegyünk fel olyan kérdéseket, amelyek negatív kicsengésűek lehetnek. Olyan kérdésekre kell szorítkoznunk, amelyek leginkább a rendszerszintű problémák feltárására irányulnak és nem az adott intézmény hiányosságait domborítják ki. Gondolnunk kell arra is, hogy amit esetleg leírnánk az intézmény kapcsán, az utólag nem hoz-e minket kellemetlen helyzetbe, kiválthat-e ellenszenvet, esetleg gátolhat-e egy későbbi jó kollegiális viszonyban, ha az adott intzménynél vállalnánk munkát.

Egy negatív példát említenék, amelyben egy államvizsgázó diák a saját munkahelyén végzett kutatást az intézmény működését illetően. Dolgozata egyik korai változatában minden interjúalanyt név és funkció szerint leírt, sőt, még az intézményben felmerült súlyos problémákról is beszámolt. Ilyen esetekben fennáll a veszélye annak, hogy a „szivárogtató" munkavállalót felettesei és kollégái ellehetetlenítik, esetleg marginalizálják és a munkahely feladására késztetik. Szerencsére még a védés és dolgozat publikussá válása előtt sikerült az alanyok identitását elfedni, kijelentéseik „élét" elvenni, a konkrét beosztásokat a „végrehajtó funkció" és a „vezető funkció" megnevezésekkel helyettesítve.

A társadalmi kutatás célja egy probléma felfedése, elemzése, semmiképpen sem egy konkrét intézmény vagy személy kritikája vagy lejáratása. Viszont az sem jó megközelítés, ha csak a pozitív dolgokról számolunk be, mert egy igényes és szakmailag kifogásolhatatlan dolgozat nem lehet intézményi „reklám". Annak itt nincs helye. Az elfogadható kutatói szemlélet értékítéletektől mentes, objektív, pro és kontra érveket ütköztet, adekvátan elemez.

2. TÉMAVÁLASZTÁS ÉS INDOKLÁS

Sokszor még a harmadév elején is határozatlanok az egyetemi hallgatók a témaválasztást és a vezető tanár személyét illetően, ezért a közelgő vezetőtanári felkérés, a téma pontosítása és a kutatási terv benyújtása időnyomás alatt történik.

Ebben a helyzetben az szoktam javasolni hallgatóimnak, hogy olyan témát válasszanak, amely kapcsolódik az eddigi szakmai gyakorlatukhoz, esetleges munkatapasztalatukhoz, munkahelyükhöz. Legyen egy konkrét elképzelésünk arról, hogy miként fogjuk bevonni kutatásunkba az alanyokat. Amint már említésre került, rendszerint könnyebb a kapcsolatfelvétel és a bizalom kiépítése azzal az intézménnyel, szervezettel, ahol esetleg önkéntes munkát végeztünk, ismerjük a munkatársakat, célcsoportot.

A témavezető tanár kiválasztásának több lehetséges szempontja is van. Az elsődleges szempont, hogy a felkérni kívánt tanár kutatási területébe illeszkedik-e a saját témánk. Másodlagos szempontok lehetnek, hogy az adott tanárnak milyenek a hallgatókkal szembeni elvárásai, kommunikációs és munkastílusa, mennyire elfoglalt, mikor vannak fogadóórái, milyen gyorsan válaszol az e-mailes megkeresésekre). Érdemesebb olyan tanárt választani, aki szívesen foglalkozik a hallgatókkal és magas elvárásai vannak, mivel a kutatás során is folyamatosan fejlődünk, tanunk és a minőségi munka jobb eredményt, osztályzatot hozhat.

Amennyiben előzetes kutatási terv leadása szükséges, tisztázni kell, hogy milyen mértékben lehetséges a tervtől való eltérés, mivel a számítani lehet arra, hogy számos olyan aspektus fog felmerülni, amely új irányt adhat a kutatásnak. Szánjuk rá az időt és úgy gondoljuk át a tervet, hogy hagyjunk magunknak némi mozgásteret a változtatásokra.

Ugyanez érvényes a címválasztásra is, a túl tág témabehatárolást általában a vezető tanár nem fogadja el, a túl szűk témabehatárolással a saját dolgunkat nehezítjük meg, amennyiben a fellelhető források vagy előzetes eredmények új irányt adnának a kutatásnak. Ennek érdekében elsősorban a földrajzi vagy intézményhez kötött behatárolást javasolnám, de akár korcsoportok vagy társadalmi helyzet szerint is pontosítható, szűkíthető a munkacím – viszont ehez tisztában kell lennünk azzal, hogy az adatgyűjtést hol és miként fogjuk végezni. Például az *„X városi fiatalok drogfogyasztási szokásai"*, *„Idősgondozás Y alapítványnál"*, *„Z egyetem hallgatóinak alkoholfogyasztási szokásai"*, *„18-22 éves korosztály párkapcsolati nehézségei"*, *„Kiskorúak online felhasználói szokásai"* stb. Az egyetem könyvtárában fellelhető és már megvédett dolgozatok témaválasztása is inspirációs forrásul szolgálhat.

A témaválasztásunkat pár sorban meg kell indokoljuk a dolgozat bevezető fejezetében, valamint ki kell hangsúlyozzuk a téma aktualitását. Az aktualitás eredhet egy társadalmi jelenség újdonság értékéből vagy a megközelítés szempontjainak újszerűségéből.

Amennyiben a hallgatónak van személyes kötődése a témához, esetleg előzetes tapasztalata, szintén érdemes erre utalni. Ilyen lehet az önkéntesség vagy munka során szerzett tapasztalat, közeli hozzátortozó érintettsége a vizsgálati probélma által stb.

A szakdolgozat megírása során a hallgatóknak az alábbi szempontoknak kell megfelelniük:

1. bizonyítsák, hogy elsajátították az önálló kutatás levezetésének képességét;

2. gyakorlatba tudják ültetni elméleti tudásukat, vagyis képesek összekapcsolni az elméletet a saját gyakorlati eredményeikkel;

3. betartják a tudományos kutatások módszetani kereteit;

4. helyesen alkalmazzák az akadémiai szerkesztésmód követelményeit.

3. A KUTATÁS MEGTERVEZÉSE

A kutatás megtervezésének számos módja és kiindulási pontja lehet. Én a kész dolgozat formailag kötött szerkezetét helyezném előtérbe, mivel a kutatási terv tulajdonképpen ennek vázlata. Amennyiben jól megírjuk a kutatás tervezetét, a későbbiekben csak ki kell bővítenünk, tehát növeljük a hatékonyságot, nem dolgozunk többször ugyanazon a problémán, és könnyebben követhetjük a haladást, illetve azonosíthatjuk a problémásabb, bonyolultabb részeket.

Az államvizsga dolgozatok szövegterjedelme általában kötött, 35-50 oldal terjedelemben határozzák meg, fedőlapok, tartalomjegyzék, könyvészeti utalások és függelékek nélkül. A túl rövid dolgozatból az tűnik ki, hogy a hallgató felületes munkát végzett és nem hozta az elvárt erőfeszítést, a túl hosszú dolgozatok arra utalnak, hogy a diák nem tud megfelelően strukturáni, vállogatni az információk között, vagyis nem volt meg a témára irányuló kellő fókusza. Ugyanez érvényes a védésre is. Amennyiben a diák később kutatói munkát fog végezni és szaklapokban szeretne publikálni, ott is kötött terjedelemmel kell számolni, ami általában 8 - 12, ritkábban 20 oldal.

Két negatív példát is felhoznék az elmúlt évek oktatói tapasztalatai alapján: az első, amikor egy hallgató dicséretre számítva, elégedetten jelentkezett egy 200 oldalas szakdolgozattal, amelyen állítólag nagyon sokat dolgozott. Ennek ellenére a dolgozat nem volt jó: túl sok általános elméletet tartalmazott, nem volt megfelelően strukturálva, a módszertani rész elnagyolt volt és az eredmények bemutatása semmilyen szempontot, kritériumot nem követett. Az ilyen „ömlesztett” munkákat kerülni kell, mivel hiábavaló a temérdek befektetett idő és elolvasott szakirodalom, ha nincs mögötte megfelelő struktúra. A mennyiségnél fontosabb a logikai strukturáltság és a felállított szempontok következetes levezetése.

A másik negatív példa, amikor egy hallgató önbizalommal jelentkezett egy terjedelmileg megfelelő, de tartalmilag „széteső” munkával. Annak ellenére, hogy gondosan megszerkesztette a küllakot, sok színes képet is bevágott, a dolgozatot az érdemi tudományos információk hiánya jellemezte. A hallgató sajnos a védés napján kénytelen volt visszalépni. A kellemetlen helyzet hátterében feltehetően az állt, hogy nem egyeztetett a témavezető tanárral, illetve nem kért elegendő segítséget.

Javasolnám, hogy amint megvan a dolgozat munkacíme, kezdjük a kötelező fejezetek szerint felépíteni és kibővíteni a dolgozatot (lásd 1. sz. Táblázat). Így összeáll a kutatási terv, majd végül a kész dolgozat.

1. sz. Táblázat – A kutatási terv javasolt szerkezete

Fejezet	Alfejezet	Magyarázat	Terjedelem
Bevezetés	A téma megjelölése és aktualitása A választás indoklása A kutatás célja	– röviden keretbe helyezzük a kutatásunkat	¼ - ½ oldal
Elméleti rész	A szakirodalom aktuális álláspontja más kutatások alapján (esetleges jogi háttér)	– általános fogalom meghatározástól kezdve a speciális tartalomig („tölcsér elv"), – ne feledjük a nemzetközi keretet is megemlíteni	kb. 10 oldal körül lesz a dolgozatban, de a kutatási tervben lehet sokkal rövidebb
Módszertan	A kutatás típusa	kvantitatív/kvalitatív esetleg vegyes	2 oldal (a kutatási tervben részletesen kidolgozzuk)
	A **célcsoportok** (kvalitatív kutatás esetén) vagy a **mintavételezés** (kvantitatív kutatás esetén)	– indokoljuk a választást a kutatás célkitűzésének függvényében, az előnyöket hangsúlyozva (mindkét módszernek vannak előnyei és hátrányai)	
	Az adott módszeren belüli kutatási eszközök	– minden esetben indokoljuk a választásunkat (a témánk vizsgálata szempontjából miért a választott eszköz a legmegfelelőbb)	
	Kutatási kérdések és hipotézis(ek)		
Gyakorlati rész	A kutatási eredmények bemutatása	– válasszunk egy fő szempontot, mely során strukturáljuk az eredményeket – ne feledjünk visszacsatolni az elmélethez	kb. 20 oldal, viszont a tervezetben nem számolhatunk be még semmilyen eredményről
Összegzés	A kutatási eredmények rövid összefoglalása		½-1 oldal, viszont a tervezetben még nincs mit összegeznünk
Bibliográfia	A felhasznált szakirodalom	– használjuk pontosan az APA hivatkozási stílust	minimális elvárás 20-30 hivatkozás a kész dolgozatban, de már a tervezetben is néhány forrást megkell említeni
Rövidítések	A dolgozatban használt rövidítések jegyzéke	– pl. eseti szám, statisztikai mutatók jelölése, egyéb rövidítések	
Függelékek	A kutatás során felhasznált eszközök, fontosabb adatok, táblázatok feltüntetése. Olyan adatok, amelyek bár fontosak, kevésbé releváns jellegük vagy a terjedelmi korlátok miatt nem kaptak helyet a szövegtestben.	– pl. az általunk használt interjú vázlata, kérdőív – saját készítésű táblázatok, amelyekből bizonyos adatokat szöveges formában említettünk a dolgozatban – Statisztikai Hivatal áltak közzétett adatok	nincs konkrét terjedelmi korlát, de arányos kell legyen a dolgozat terjedelmével (javasolt kb. 5-15 oldalt nem meghaladni)

A fenti táblázatból kitűnik, hogy 1/3 elmélettel és 2/3 gyakorlati, azaz empirikus résszel ajánlatos tervezni. Régebben elfogadott volt a 40% elméleti, 60% empirikus kutatási arány is.[3] A táblázatban foglalt terjedelmi arányoktól az adott egyetem vagy szakirány eltérő követelményeket is támaszthat, tehát előzetesen tájékozódjunk a ránk vonatkozó formai és tartalmi követelményekről.

Amennyiben túl kiválasztottuk és egyeztettük a témánkat a vezető tanárral, az első lépés még a kutatási terv kibővítése előtt a szakirodalom előzetes áttanulmányozása. Fontos, hogy kezdjünk el szakirodalmat olvasni, mielőtt véglegesítenénk a kutatási tervünket. Így rálátásunk nyílik arra, hogy mik az aktuális kérdéskörök egy adott területen és ennek megfelelően fókuszáljuk a témánkat.

Határozottan ellenjavallt, hogy a saját fejünk után kezdjük behatárolni a kutatási irányokat, vagy interjúvázlatot, esetleg kérdőívet szerkeszteni! Bár lehet az elején egyszerűbbnek tűnik a saját logikánkra és elképzeléseinkre hagyatkozni, mintsem rászánni az időt az olvasásra, de a későbbiekben emiatt szinte biztosra vehető, hogy nagy gondban leszünk és nem fog „összeállni" a kutatás. Az előzetes szakirodalmi olvasmányaink abban segítenek, hogy eleve egy viszonyítási alappal indulunk neki a kutatásnak, ismerjük, hogy más kutatók milyen eredményekre jutottak egy adott témában, így a saját kutatási kérdeseinket ezekre az eredményekre alapozzuk, ahogyan az interjúvázlatot vagy kérdőívet is ezen eredmények és kutatási irányok fogják isnpirálni. Ezáltal nagymértékben megkönnyítjük saját magunk számára az eredmények visszacsatolását az elmélethez.

Bár az előbbiekben utaltunk arra, hogy jó, ha a kutatásunk rendelkezik bizonyos újdonság értékkel, ez nem azt jelenti, hogy teljesen újat kell kitalálnunk, hanem azt, hogy egy már „kitaposott ösvényen" elindulunk, vagy ezeket kombinálva a már „meglévő elemekkel" építkezünk. Tehát határozott elvárás, hogy egy adott területen a már meglévő információkra, eredményekre alapozzunk a saját kutatásunkban.

3 A szakcikkek esetén másak az arányok, az elméleti rész sokkal rövidebb, mivel szakemberek fogják olvasni, akik feltehetően ismerik a szakirodalmat.

4. A KUTATÁS ELMÉLETI MEGALAPOZÁSA

A kutatás ötletének megszületése után feltérképezzük a témánkra vonatkozó szakirodalom aktuális álláspontját. Ez segít minket a tervezésben, a célkitűzésünk megfogalmazásában és a kutatás irányának meghatározásában. Ebben a fejezetben arra térnék ki, hogy – függetlenül a választott témától – milyen információkat és hogyan szerezzünk be, hol keressük a szakirodalmi forrásokat és mit jegyezzünk fel már az első olvasás során.

Néhány alapkövetelményt érdemes szem előtt tartanunk:

1. Keressünk minél újabb forrásokat, mert ezek tükrözik a témánk aktuális helyzetét, illetve az eddigi kutatási eredményeket. Ajánlott, hogy a hivatkozott forrásokat – a társadalomkutatás gyors fejlődése miatt – lehetőleg az utolsó 1-5 év publikációiból válogassuk össze. Természetesen használhatunk régebbi forrásokat is, különösen ha egy adott terület alapműveiről van szó, vagy amikor történelmi hátteret vizsgálunk, esetleg fogalommeghatározást végzünk. Gondoljunk csak bele, hogy az online közösségi média használatának szokásai milyen gyorsan változnak, vagy akár az elmúlt évek társadalmi változásai mint pl. a Covid-19 mekkora változásokat hozott.

2. A nemzetközi szakirodalom nem megkerülhető, alapos munkához nemcsak a hazai kutatásokat, elméleti munkákat, hanem a külföldi vonatkozásokat is fel kell használnunk.

A legtöbb nívós szakcikk elsősorban angol nyelven érhető el, tehát az angol nyelvtudás nagymértékben megkönnyíti a dokumentálódást.

3. Használjuk az egyetemi könyvtárban elérhető online adatbázisokat, mert a leghatékonyabban így tudunk a témánkba vágó kulcsszavak után keresni. A Google Tudós is sokat segíthet, de mindenképpen kerüljük az olyan forrásokat, amelyek nem hitelesek, a szerzők nem ismertek vagy nem szakemberek, mint pl. a Wikipedia vagy különféle blogok, influenszerek honlapjai. Az utóbbi években tapasztaltam, hogy a hallgatók egyre gyakrabban hivatkoznak népszerű youtuberekre vagy ezek csatornáira bizonyos közéleti témák kapcsán, ez szintén kerülendő, mert nem tudományos.

Keressünk kulcsszavak után, majd a találatokat szűrjük a kivonatok vagy absztraktok áttekintésével, mivel a legtöbb találat nem lesz számunkra releváns.

4. Már az eslő áttekintés során a számunkra releváns forrásokból kezdjük kiemelni (copy-paste) egy külön dokumentumba azokat a mondatokat, amelyeket lehet, hogy fel fogunk használni a szerkesztéshez. Utólag még mindig könnyebb kitötölni, amit nem használunk, mint ismét kikeresni, illetve átszerkeszthetjük,

strukturálhatjuk. Viszont már abban a pillanatban, amikor átmásoltunk egy mondatot, jegyezzük fel gondosan a hivatkozást és az oldalszámot is, hogy még véletlenül se essünk a plagizlás hibájába! A legtöbb adatbázis és online szaklap eleve támogatja a hivatkozás APA stílusban történő kigenerálását.

Az így kimásolt információ fogja a dolgozatunk elméleti részét megalapozni, akár szó szerint idézve, akár tartalmilag hivatkozva. Nekünk főként a lejegyzett mondatokból kell szelektálnunk, strukturálnunk és az összekötő szöveget kell megírnunk, viszont nem győzőm eléggé hangsúlyozni, hogy mindig pontosan hivatkozzunk a szerzőre, forrásra, akitől/ahonnan bármit átvettünk.

5. Semmi nem nyilvánvaló és köztudott a tudományos munkákban, így a szakdolgozatokban sem az. A személyes véleménynek, értékítéletnek nincs helye, mivel minden kijelentésünket objektíven tényekkel és eredményekkel alá kell támasztanunk. Minden egyes kijelentésünkre keressünk megfelelő forrást, szakcikket, könyvet, statisztikai adatot. Itt nem célunk, hogy eredetiek és kreatívak legyünk, hanem alaposan dokumentáltak. Ennek ellenére mégis lesz általunk hozzáadott érték, mely a szerkezeti felépítésben, struktúrában és stílusban lesz tetten érhető. A dolgozat elméleti részében a hallgató saját hozzájárulása a fellelhető információk logikus strukturálása; a hallgató saját magától szinte csak összekötő mondatokat kell írjon. A hallgató viszont lehet kreatív az eredmények bemutatásánál, amikor magyarázza és értelmezi a saját kutatási eredményeit.-

6. Ha úgy tűnik, hogy a választott témánknak nincs bőséges szakirodalma, akkor biztosra vehetjük, hogy nem kerestük eléggé alaposan. Szinte bizonyos, hogy nincsenek olyan témák, amelyeket előttünk számtalan kutató ne dolgozott volna már fel. Keressünk tovább és találni fogunk! Az alapos dokumentálódás ismérve a többnyelvű könyvészeti lajstrom – amire a 2-es pontnál már utaltunk.

5. ONLINE ADATBÁZISOK

Időkihasználás, hatékonyság szempontjából a leghasznosabb az online adatbázisokban való kutatási terület, ezen belül a kulcsszavak alapján történő keresés. Az online adatbázisok előnye a Google kereséssel szemben, hogy nemcsak az ingyenesen elérhető források jeleníthetők meg, hanem azok a tudományos munkák is, amelyek megtekintése fizetéshez kötött. Otthonról erre nincs lehetőségünk, de az egyetemi könyvtárak (és a városiak is) rendelkeznek némely adatbázisért fizetett bérlettel. Szerencsés esetben olyan könyvtári belépőt is kaphatunk, hogy a könyvtár honlapjáról a saját belépőnkkel otthonról is hozzáférhetőek lesznek az adatbázisok; erre a lehetőségre érdemes rákérdeznünk könyvtárosunknál.

Ilyen adatbázisok – a teljesség igénye nélkül (rendszerint angol nyelven):

1.	JSTOR	*https://www.jstor.org/*
2.	SpringerLink	*https://link.springer.com/*
3.	EBSCO	*https://www.ebsco.com/*
4.	ResearchGate	*https://www.researchgate.net/*
5.	ProQuest	*www.proquest.com/*
6.	Scopus	*https://www.scopus.com/*

Ha problémás megoldanunk a könyvtári tartózkodást, akkor segítségünkre lehet a Google Tudósban (https://scholar.google.com/) indított keresés, de például a ResearchGate honlapjáról is számos full-text szakcikk elérhető (https://www.researchgate.net/).

Magyar nyelvű adatbázisokkal is rendelkezünk, bár szűkebb témakörökben lesz eredményes a keresésünk:

1.	L'Harmattan Digitális Adatbázis	*https://www.szaktars.hu/harmattan/*
2.	Osiris Kiadó	*https://www.szaktars.hu/osiris/*
3.	Akadémiai Kiadó MERSZ	*http://mersz.hu/*
4.	Kronosz Kiadó	*https://www.szaktars.hu/kronosz/*
5.	Szaktudás Kiadó	*https://www.szaktars.hu/szaktudas/*
6.	Typotex Interköny	*https://edu.interkonyv.hu/*
7.	Arcanum Digitális Tudománytár (ADT)	*https://adtplus.arcanum.hu/hu/*
8.	Akadémiai Kiadó Szótárai	*http://www.szotar.net/*
9.	SzóTudásTár	*https://szotudastar.hu/*
10.	Mentor Kiadó	*https://adtplus.arcanum.hu/hu/collection/KiadokMentor/*

Forrás: Partiumi Keresztény Egyetem, é.n.

Feltételezve, hogy van hozzáférésünk, honnan is kezdjük a keresést? Mivel egy adott szakcikk általában több adatbázisban is hozzáférhető, ezért kevésbé lényeges, hogy melyik adatbázist választjuk. Nyilvánvaló, hogy több adatbázis szimultán használata növeli a találati esélyeinket.

Az első lépésben írjunk össze néhány kulcsszót, mely szorosan kötődik a választott témánkhoz, majd módszeresen kezdjünk keresni az adatbázisban. Minden kulcsszavunkhoz rendelhetünk egy külön mappát, ahová a találatokat fogjuk összegyűjteni, majd külön gyűjtő dokumentumba elmenthetjük a tartalom (kigenerált vagy kimásolt) bibliográfiai adatait (mivel az elmentett szövegtestben nem mindig vannak benne). Így a későbbiek során kéznél lesznek és tudunk belőlük dolgozni. Azokat a találatokat érdemes megnyitanunk, amelyek címe a témánkba vág és viszonylag friss tanulmányok. A következő lépés a kivonatok (absztrakt) áttekintése és a tartalom lementése.

A cikkek teljes terjedelemben való átolvasását a kivonatok és konklúziók számunkra releváns jellege alapján szűrhetjük. A számunkra releváns információkat, amiről úgy gondoljuk, hogy helye lehet a dolgozatunkban, már az első olvasáskor érdemes elmenteni – hivatkozással együtt, ahogyan az előző fejezetben már utaltunk rá.

Véleményem szerint jobb időhatékonyságot érhetünk el, ha a keresési fázist, a források eltárolását, majd ezek átolvasását különválasztjuk. Így az elméleti megalapozás megszerkesztésének egy részét akár offline is elvégezhetjük, feltéve, ha közben nincs szükségünk egy online fordító programra.

Az általam javasolt munkamódszerrel egy többoldalas, akár többnyelvű „ömlesztett" piszkozatunk lesz, tele hivatkozásokkal. A továbbiakban szintetizálni és strukturálni kell. Olvassuk újra a kijegyzetelt gondolatokat és idézeteket, majd próbáljunk egy logikai fonalat elképzelni, hogy mely információk csoportosíthatóak, illetve melyik gondolatokat tudnánk minimális összekötő szöveggel átvezetni egy másikra. A későbbiekben még szükséges lehet az elmélet kiegészítése, sőt átstrukturálása is, de ezáltal nagyvonalakban teljesítettük a szakdolgozat elméleti megalapozását.

A fent leírt munkamódszer mindössze egy javaslat, amely véleményem szerint rendkívül hatékony. Természetesen másképpen is meg lehet oldani a feladatot, az egyénileg kialakított preferenciák és rutinok szerint.

6. HOGYAN HIVATKOZZUNK?

A hivatkozási szabályok szakmai egyezmények eredményei, mely során egységesen kodifikálták az adott tudományterületen alkalmazandó bibliográfiára vonatkozó formai követelményeket. Számos hivatkozási stílus létezik, ennek megfelelően könnyen hozzáférhető hivatkozási útmutatókat tölthetünk le az internetről. A hivatkozási útmutatók az egyetemi honlapokon is elérhetőek. A hivatkozások fő funkciója a szellemi tulajdonjog védelme, a dolgozatunkba átemelt gondolatok, idézetek és adatok forrásának megjelölése által.

Mivel az egyetemi tanulmányok első félévétől kezdődően minden szakdolgozathoz, bemutatóhoz és egyéni vagy csoportos referátumhoz következetesen elvárt követelmény a hivatkozási szabályok betartása, feltételezhető, hogy a hallgatóknak nem jelent különösebb nehézséget ezek alkalmazása a szakdolgozatban. Azonban vannak kivételek, tehát pár szót – a teljesség igénye nélkül – erről a témáról is ejtenék.

A szociális munkában, szociológiában, antropológiában és pszichológiában az általánosan elfogadott hivatkozási stílus az American Psychological Association nevéhez fűződik, melyet röviden APA hivatkozási stílusnak neveznek. A szabványosított hivatkozás eredete 1929-re nyúlik vissza, egy szerző-dátum alapú rendszer; jelenleg aktuális 7. Kiadványa a *Publication Manual of the American Psychological Association.* Hivatkozási minták például a https://apastyle.apa.org/style-grammar-guidelines/references/examples honlapon is elérhetőek.

A hivatkozási stílus használata két lépésben történik:

1. A szövegtestben a szerző nevét és évszámot írjuk kerek zárójelben, amennyiben gondalatot, összefoglalást veszünk át, így: (Albert-Lőrincz, 2018), viszont ha parafrazálunk vagy szó szerint idézünk, az oldalszám megjelölése is szükséges, pl.: (Albert-Lőrincz, 2018, 19.).

2. A dolgozat végén található hivatkozási listánkba a hivatkozási forrás típusának megfelelő formátumot használjuk. A forrás típusa sokféle lehet: szakkönyv, szakcikk, újságcikk, intézményi honlap, szerző nélküli honlap, film, blog, videó stb. Az irodalomjegyzék tetszőlegesen nevezhető könyvészetnek, hivatkozásoknak vagy bibliográfiának.

A gyakran használt forrástípusok hivatkozása akár készséggé is fejlődhet, viszont a ritkábban használt források formátumát mindig ellenőrizzük valamely útmutatóban, vagy akár online példákra is rákereshetünk.

Lehetőleg mindig az elsődleges forrásokra hivatkozzunk, ami azt jelenti, hogy amikor utalást találunk egy szakcikkben más szerzőkre, tanulmányokra, keressük vissza az eredeti forrást. Ez logikus követelménye az információ korrekt átadásának, mivel egy szakcikkben már fellehető hivatkozás tartalmazhatja az adott szerző szempontja szerint kiemelteket vagy sajátos értelmezését.

Visszatérő hiba a dolgozatokban, hogy a diák csak az URL-t (Web címet) adja meg; ez sosem elegendő, egyrészt mert nem tartja be az előírt formátumot, másrészt a Web címek nem állandóak, előfordulhat, hogy idővel megváltozik a honlap tartalma, törlik, vagy áthelyezik azt.

APA hivatkozási stílusban a táblázatokban csak vízszintes vonalak szerepelnek, a táblázatot felül számozzuk és a szövegtestben minden egyes táblázatra szöveges formában is utalnunk kell. Az adatvizualizáció tehát bizonyos kötöttségekkel jár. A 2. sz. Táblázat a fentieket példázza:

2. sz. Táblázat – Mennyire elégedett az egyetemi képzéssel?

		Gyakoriság	Százalék	Érvényes százalék
Érvényes válaszok	Egyáltalán nem	21	9.0	10.7
	Kismértékben	11	4.7	5.6
	Közepesen	83	35.5	42.3
	Jól	65	27.8	33.2
	Teljesen jó	16	6.8	8.2
	Összesen	196	83.8	100.0
Hiányzó válaszok		38	16.2	
Összesen		234	100.0	

(Forrás: Albert-Lőrincz, Belényi, 2019, 5.)

Az ábrákat az ábra alatt számozzuk; minden egyes ábrára utalni kell a szövegtestben. Az olvashatóság érdekében az ábrára ki kell tenni a pontos százalékos vagy esetszámos értékeket.

Itt egy példa ábrára:



1. sz. Ábra – Egyetemi képzés színvonala

(Forrás: Albert-Lőrincz, Belényi, 2019, 6.)

A táblázatokhoz és ábrákhoz minden esetben rövid magyarázatot kell fűzni. Két táblázat vagy ábra között feltétlenül szövegnek is kell lennie. Fejezetet vagy alfejezetet nem elfogadott táblázattal vagy ábrával zárni, mivel a táblázat vagy ábra után magyarázó szöveget kell adni. Amennyiben átvettünk szerzőtől vagy Statisztikai Hivataltól egy táblázatot vagy ábrát, ne feledjük zárójelben a forrást is megjelölni.

A táblázatokat és az ábrákat külön-külön számozzuk (adott dolgozatban 1. sz. Táblázat, 2. sz. Táblázat stb., ugyanígy 1. sz. Ábra, 2. sz. Ábra stb.).

A hivatkozások pontos használata időigényes, tehát szánjunk rá türelmet, mert elemi formai követelménye a szakdolgozatoknak.

A Word szövegszerkesztőben található egy beépített hivatkozási funkció (References>Insert Citation), ennek használatát csak nagy körültekintéssel ajánlanám, mivel használata megszokást igényel. A hivatkozás összes adatát amúgy is meg kell adni, majd egy kiegészítő lépésben hozzárendelni az oldalszámot, amennyiben szó szerint idéztünk (Edit citation). A szakdolgozat 15-25 hivatkozása véleményem szerint manuálisan is könnyedén kezelhető. A hivatkozási funkciónak kétségtelenül vannak előnyei is; például akkor érdemes használni, ha nagyon sok hivatkozásunk van (doktori disszertációk esetén akár 150-200), vagy, ha a dolgozatunkat külön dokumentumokból részenként rakjuk össze, esetleg részekre szednénk szét, vagy, ha váltani szeretnénk a hivatkozási stílusok között az egységesítésük céljából.

Amennyiben a hivatkozásainkat manuálisan kezeljük, figyeljünk arra, hogy a szerzőket névsor szerint tüntessük fel a dolgozat végén, az APA hivatkozási követelménye szerint, tehát nem előfordulási sorrendben; sorszámozásuk nem szükséges. A hivatkozásokban törekedjünk egységesen használni a pontokat és/vagy veszőket a megfelelő külalak kialakítása érdekében.

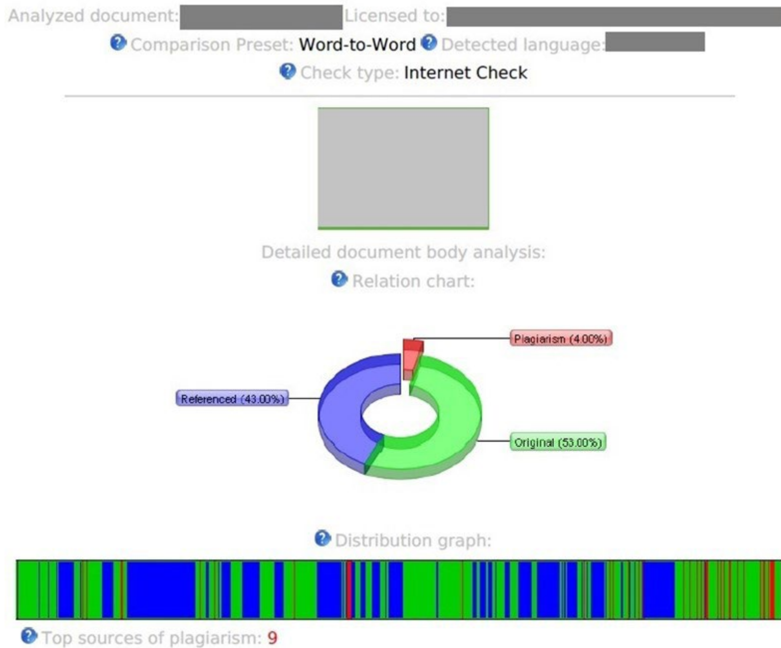
Hasznos linkek:
APA Style honlapja: https://apastyle.apa.org/
Hivatkozási útmutató: HIVATKOZZ HELYESEN! (é.n.), http://padi.psiedu.ubbcluj.ro/kolozsvar/wp-content/uploads/Hivatkozasi-utmutato-BA.pdf (hozzáférés 2019.07.17-én)
Hivatkozási példák: Pennstate University Libraries, APA Quick Citation Guide (é.n.) https://guides.libraries.psu.edu/apaquickguide/otherformats (hozzáférés 2019.07.17-én)

7. MIT JELENT A PLAGIZÁLÁS ÉS HOGYAN ELLENŐRZIK?

A plágium, plagizálás szavak latin eredetűek, a szerzői tulajdonjog megsértésére utalnak. Valaki akkor plagizál, amikor más munkáját, gondolatait forrásmegjelölés vagy szerzői engedély nélkül a sajátjaként tünteti fel. A saját fordítású szövegeknél is meg kell jelölni az eredeti, idegen nyelvű forrást, a szerző nevét, a dokumentum címét és lelőhelyét. Akadémiai körökben súlyos vétségnek számít, akár a megszerzett diplomák visszavonását is eredményezheti.

Az elmúlt években a különféle szoftverek elterjedésével sokat változott a plágiumellenőrzés. A 2010-es évek elején alkalmazott plágiumellenőrzési folyamat ma már megmosolyogtató módon történt (például az egyetemeken doktoranduszoknak adták ki feladatként, hogy a dolgozatokat a Google keresőben ellenőrizzék). A dolgozatból mondatokat, vagy rövidebb szövegrészeket kellett copy-paste-tel a keresőbe beilleszteni, és ahol forrás megjelölés nélküli egyezést találtak, azt jelezni kellett. Ez a módszer természetesen nem volt sem hatékony, sem pontos, viszont mára már szerencsére vannak sokkal jobb módszerek.

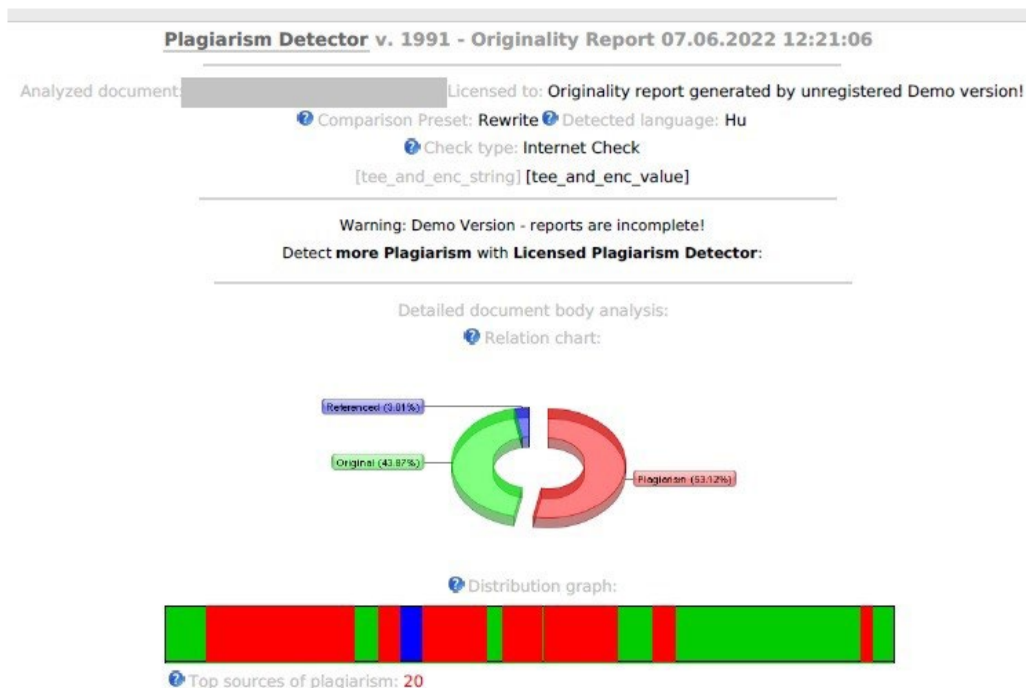
A plágiumellenőrző szoftverek is hasonlóan működnek, csak sokkal alaposabban; nem csak a Google keresőben ellenőrzik az esetleges egyezéseket, hanem különböző adatbázisokkal is összevetik a szöveget. Ez a módszer sem tökéletes, mert az interneten nem fellelhető forrásokkal nem tud mit kezdeni; a szoftveres ellenőrzés olyan részeket is megjelölhet, amelyek nem „lopottak", csak egyezést mutatnak valamivel. Gyakran előfordul, hogy a törvényszövegeket vagy intézményi honlapokról származó információkat plágiumnak tekinti, így a dolgozat bizonyos (minimális) százaléka a jelentés szerint plagizált lesz. Ez a bizonyos százalék kisebb kell legyen az adott egyetemen, illetve képzésen, szakirányon elfogadhatónak tartott maximumnál, ami általában 5-10% között mozog. Az alábbi 2. sz. Ábra egy plágiumellenőrzés eredményének jelentése:



2. sz. Ábra – Licenszelt plágiumjelentés

Az fenti plágium jelentés egy államvizsga dolgozatról készült. A jelentés szerint a dolgozat 4%-a plagizált, 43%-a hivatkozott és 53%-a eredeti, ami teljes mértékben elfogadható.

Az alábbiakban egy olyan jelentést láthatunk (3. sz. Ábra), amelyet egy félévi referátumon futtattam le. Feltűnt, hogy egy másodéves diák szintjén szokatlanul jó a dolgozat, túl szakmai és szabatos stílusban íródott. Mint kiderült, a diák a dolgozata több mint felét másolta. A plágiumellenőrzés eredménye:



3. sz. Ábra – Démó plágiumjelentés

Érthető, hogy nagy a kísértés a másolásra, plagizálásra, mert számos témában kész referátumok tömkelege tölthető le az internetről. Szerencsére az érintett diák még idejében tanulhatott a hibájából és harmad évben egy korrektül megírt szakdolgozatot mutatott be.

A két bemutatott plágiumellenőrzés között az a különbség, hogy az első példánál egy intézményi licenszelt jelentést, a másodiknál egy demó verziós jelentést láthatunk. A demó verziónál fennállhat oldalszám korlátozás, vagy az ellenőrzés mélységének korlátozása, viszont így is használhatjuk, akár otthon is, dolgozatunk részekre bontott ellenőrzésére.

Az interneten plágiumellenőrző szoftver címszó alatt nagyon sok termék közül válogathatunk. Bármelyik szoftver használható, mivel hasonló a működési algoritmusuk, de általános érvénnyel kijelenthető, hogy amiért fizetni kell, jobban fog teljesíteni, mint az ingyenes verziók. Amennyiben lelkiismeretesen dolgoztunk és minden forrást megjelöltünk, nem lesz gond a szoftveres plágiumellenőrzés során. Sokkal könnyebb erre menet közben figyelni, mint utólag a hibákat javítgatni és az elmúlasztott hivatkozásokat pótlólagosan kikeresni. Az otthoni előzetes plágiumellenőrzést a szakdolgozat leadása előtt akkor javasolnám, ha bizonytalanok vagyunk a munkánk ezen részével kapcsolatban.

8. A MESTERSÉGES INTELLIGENCIA NAGY NYELVI MODELLJEINEK HASZNÁLATA

A mesterséges intelligencia napjainkra nemhogy rohamos fejlődésen ment keresztűl, hanem bizonyos típusai, mint például a nagy nyelvi modellek bárki számára elérhetővé váltak. 2022 november 30.-án az OpenAI nyilvános használatra bocsátotta a ChatGPT chatbotot, amely rekord idejű felfutása felhasználók százmillióit vonzotta különböző feladatok elvégzésére. Köztük középiskolás diákok és egyetemisták, sőt tanárok és kutatók is szép számmal akadtak, akik iskolai vagy tudományos-akadémiai feladataikat hárították a chatbotra. Jelen fejezetben csak érinteném a jelenséget, mivel a téma mélységeiben szinte kimeríthetetlen.

A jelenség annyira új, hogy véleményem szerint jelenleg erre még nincs felkészülve az oktatási rendszer, bár ennek veszélyei már most nyilvánvalóak: az iskolai és egyetemi feladatok a diákok tudását és készségeit hivatottak fejleszteni, így aki ezt másra, vagy akár a mesterséges intelligenciára bízza, nem fog fejlődni. Továbbá nem etikus ennek használata azon diáktársakkal szemben sem, akik maguk oldják meg a feladataikat “mankók” és külső segítség nélkül. A mesterséges intelligencia iskolai és egyetemi feladatokhoz való felhasználása megnehezíti a valós teljesítmény alapján történő értékelést, rangsorolását, esetenként tanulmányi ösztöndíjra jogosultság megállapítását.

A kutatók közül is sokan felhasználják a mesterséges intelligencia lehetőségeit publikációikhoz, ezáltal jelentős időt megtakarítva versenyelőnybe hozva magukat a hagyományos módszerekkel kutató és publikáló társaikkal és kollégáikkal szemben. Bár a színvonalas szaklapok ma már elviekben nem fogadnak be chatbotok segítségével írt publikációkat, ezek szaklektorok általi kiszűrése, ellenőrzése szinte lehetetlen.

Ma már a legtöbb plágiumellenőrző szoftver rendelkezik mesterséges intelligencia detektorral is, viszont ez a technológia tapasztalataim alapján megbízhatatlan: olyan esetekben nem ismeri fel a gépi szöveget, amikor ismert a nyelvi modell használatának ténye, illetve olyan publikációk esetében is képes gépi szöveget felismerni, amelyek még a nagy nyelvi modellek technológiája előtt íródtak.

Az egyetemi képzésben át kell gondolni az otthoni referátumok alkalmazását, mert jelenleg ellenőrizhetetlen, hogy maga a hallgató, vagy esetleg egy nagy nyelvi modell oldotta meg a feladatot.

A hallgatóimnak azt tanácsolnám, hogy tanuljanak meg élni a mesterséges intelligencia adta lehetőségekkel, tartsanak lépést korunk technikai vívmányaival munkaerőpiaci versenyképességük megőrzésének érdekében. Azonban

lelkiismeretükre bíznám, hogy ne éljenek vissza ezen lehetőséggel az iskolai/egyetemi képzés során. Tekintsék a képzést egy lehetőségnek, ami nemcsak tudásukban, készségeikben, hanem személyiségükben is fejleszti őket, feltéve, ha a saját képességeikre hagyatkoznak.

Tapasztalataim szerint a nagy nyelvi modellek által generált szakszövegek egyik jellemző gyenge pontja a hivatkozásokkal kapcsolatos, ugyanis gyakran nem létező, vagy nem pontos a megadott referencia. További jellemzője ezen szövegeknek, hogy gyakran túl általános jellegűek. Előfordulhat egy kivonat készítése során, hogy olyan elemeket tartalmaz a gépileg generált szöveg, amelyek nem képezték az eredeti szöveg részét, releváns részleteket is kihagyhat, valamint a szavak eredeti jelentését is „átköltheti".

Tehát a mesterséges intelligencia jelenlegi fejlettségi szintje nem alkalmas minőségi szakszövegek generálására, mivel tartalmilag gyakran hibázik. Azonban bizonyos mechanikus részfeladatokban segítségünkre lehet, mint például egy interjú hangfelvételének szöveges átirattá alakításában, vagy ezen szövegek egyenes (direkt) beszédből függő (indirekt) beszéddé történő alakításában. Kérhetünk szövegösszefoglaló tartalmakat a nagy nyelvi modellektől, sőt használhatjuk akár tartalmi elemzésre is. A megfelelő bővítmények telepítése után többek között lehetőségünk nyílik külső webhelyek, URL-k elérésére, sőt bármilyen PDF-et is feltölthetünk elemzés céljából. Azonban mindig ellenőriznünk kell, hogy a kapott eredmény megfelel-e az eredeti szövegünk tartalmának.

Összességében úgy értékelném a mesterséges intelligencia adta technológia lehetőségeket a társadalomtudományi kutatások szempontjából, hogy körültekintően használva megkönnyítheti a kutatói munkák bizonyos mechanikus részeit, viszont abban ne bízzunk, hogy helyettünk megfelelő minőségben képes lenne elvégezni a teljes feladatot.

9. MITŐL VÁLIK TUDOMÁNYOS JELLEGŰVÉ EGY ÉRTEKEZÉS?

A pontos hivatkozások mellet, amint már utaltam rá a 4. fejezet 5. pontjában, elsődleges az alapos dokumentáltság, a tények objektív bemutatása eredmények formájában. Javasolnám, hogy minden kijelentésünk és megállapításunk után tegyük fel a kérdést magunkban, hogy „ezt honnan is tudjuk?".

Az eredmények logikus értelmezése is hozzájárul a kutatás értékéhez, ezek megléte arra utal, hogy az illető diák plusz erőfeszítést hoz azért, hogy a kapott eredményeket társadalmi összefüggésekben interpretálja.

Az objektivitás az is jelenti, hogy amennyiben hátrányokról beszélünk, említsük meg az előnyöket vagy potenciális lehetőségeket is. Ütköztessük a pro és kontra érveket, a saját értékítéletünket pedig mellőzzük. A gyakorlatban ez nehézséget jelenthet, mivel a hallgató még nem feltétlenül rendelkezik a szükséges szisztematikus tudással, viszont ezen szándékának és törekvésének nincs akadálya.

A kutatásunk hozhat akár a szakirodalommal ellentétes eredményeket is; a kutatási kérdésinkre lehet nemleges válasz is; hipotéziseinknek nem feltétlenül szükséges „beigazolódniuk", mivel a negatív eredmények is ugyanolyan értékesek, mint a pozitív eredmények.

Ebben az esetben próbáljunk magyarázatot találni az eltérésre: lehet, hogy, más volt a vizsgálati szempontunk, az is lehet, hogy időben megváltozott a társadalmi jelenség. Példaként említeném az online világot, ahol egyik évről a másikra teljesen megváltozhatnak a felhasználói szokások, például egy új applikációnak köszönhetően.

A nyelvezetet tekintve mellőzzük a jelzők használatát, tehát kerülendő az eredményeink jelzők általi kicsinyítése vagy felnagyítása. A saját véleményt is mellőzzük, mindig vagy a szakirodalomra, vagy saját adatainkra alapozzunk.

Igyekezzünk választékosan fogalmazni, kerüljük a szóismétléseket, törekedjünk az egyértelmű és világos megfogalmazásokra. A dolgozat újraolvasása során a stílus nagymértékben javítható. Jellemző hiba az is, hogy a hallgató leírja, hogy bizonyítani, igazolni szeretné a felvetését. A helyes megközelítés a felvetés, hipotézis vagy kutatási kérdés vizsgálata, függetlenül attól, hogy végül igazoljuk vagy elvetjük azt.

10. A DOLGOZAT STRUKTURÁLÁSÁNAK FONTOSSÁGA

Az ebben a fejezetben leírtak érvényesek lehetnek úgy az elméleti, mind a gyakorlati rész vonatkozásában. Egy jól strukturált dolgozat olyan, aminek a tartalomjegyzékéből is világosan kitűnnek a gondolati egységek, a kutatás egészében fellelhető logikai fonal, amely a célkitűzésnek van alárendelve. Ennek megértése nem egyszerű, ezért néhány támponttal próbálok segíteni a szakdolgozat készítőinek.

A dolgozat strukturálása egy folyamatos tevékenység, ahol a „végső simítások" akkor tudnak megvalósulni, amikor tartalmilag minden összeállt. Ugyanaz a tartalom eredményezhet jó és rossz dolgozatot is az adatok bemutatásának sorrendjétől, struktúrájától függően. A jó és átgondolt szerkezeti struktúra emeli a szakdolgozat minőségét.

Íme néhány javaslat, amely segíthet a rendszerezésben:

1. A téma aktualitásának és a kutatás célkitűzésének pontos meghatározása.

2. Az elmélet következetesen alátámasztja a célkitűzésben megfogalmazottakat.

3. Az elmélet indulhat a vizsgált fogalmak vagy jelenség meghatározásával. A meghatározott fogalmat általános társadalmi kontextusba kell helyezni. Az általános kontextus feltehetően nemzetközi, európai uniós és nemzeti szinten is fellelhető. Innen levezethető a téma függvényében választott kutatási célcsoport és a hozzájuk kapcsolódó jellemzők, nehézségek (ebben az olvasott szakirodalom lesz segítségünkre). Amennyiben a téma indokolja, a jogi háttérre is ki lehet térni. Ha egy adott intézményen belül kerül sor a kutatásra, az intézmény tevékenységét is be lehet mutatni röviden. Javasolt tehát az általánostól a különleges felé haladni, a „tölcsér elv" szerint.

4. Az elmélet és a gyakorlat összefügg. Ez alatt azt értjük, hogy a szakirodalomban fellelhető problémafelvetést tükröznie kell a módszertanunk kutatási kérdésekkel és hipotézisekkel foglalkozó részének. A gyakorlatban a saját eredményeinket visszacsatoljuk, azaz összehasonlítjuk az elméleti részben felvetettekkel.

5. Az eredmények bemutatásának sorrendje nem az alkalmazott módszerek sorrendjével kell megegyezzen, hanem inkább a kutatási kérdéseink sorrendjével. Így a kutatási kérdéseink sorrendje sem mellékes, egyik kérdésből következnie kell, vagy legalább a kérdéshez kapcsolódnia kell a következő kérdésnek.

6. Javasolt több alcím beiktatása is, de az alcímek között kapcsolat kell fennáljon. Ezt úgy érhetjük el, hogy az adott alcím végén összegzés formájában kivezetjük a témát egy-két mondatban, majd a következő alcímben utalunk az előző alcímben foglaltakra és ehhez kötjük az új aspektusokat.

7. Egy adott gondolatot csak egy helyen említünk, nem szükséges ismételnünk önmagunkat. Ez hatványozottan érvényes az eredmények bemutatására is.

Egy dolgozat megítélését nagymértékben befolyásolja annak struktúrája, mivel egy rosszul strukturált dolgozat arra enged következtetni, hogy a hallgató nincs tisztában a célkitűzésével, vagy hanyagul állt a feladathoz.

Amikor értékelő tanárként kézbe veszek egy szakdolgozatot, először fellapozom a dolgozatot és elolvasom a kutatás módszertani leírását, fejezetét. Ebből kitűnik, hogy a hallgatónak mi volt a kutatással kapcsolatos koncepciója, valamint miként folytatta le a kutatást. Következő lépésben a dolgozat végére lapozok és átnézem a bibliográfiai hivatkozásokat. Ebből ránézésre látszik a szakirodalom minősége, mennyisége, aktualitása, valamint utal az akadémiai szerkesztési szabályok ismeretére és betartására. A harmadik lépésként ismét a dolgozat elejére lapozok és megtekintem a dolgozat szerkezetét, felépítésének logikáját a tartalomjegyzék alapján. Negyedik lépésként ismét hátra lapozok, és szúrópróba szerűen beleolvasok az eredmények bemutatásába. Pár mondat alapján kiderül, hogy milyen ennek a színvonala és a hallgató megkísérelte-e az eredméneyk értelmezését. A dolgozatról ez által a négy lépés által kialakul egy benyomásom, amit csak ritkán változtat meg annak konkrét, részletes tartalma.

11. KUTATÁSI MÓDSZEREK ÉS ESZKÖZÖK

Jelen fejezetben a kutatási módszerekről és eszközökről lesz szó. A kutatási módszer egy bizonyos eljárásra, mely az adatgyűjtés és feldolgozásá mikéntjére, technikájára utal. Többféle osztályozás létezik, ezek közül megemlítenénk a primer és szekunder, valamint a kvantitatív és kvalitatív módszereket.

A primer kutatás során magunk végezzük az adatgyűjtést és feldolgozást minőségi és mennyiségi módszerekkel (mint például, kérdőívezés, interjúk, fókuszcsoport, terepkutatás-megfigyelés), a szekunder kutatás során a már meglévő adatok elemzését végezzük (mint pl. Statisztikai Hivatal adatai, korábbi kutatások adatainak új szempontok szerinti értelmezése). A primer kutatásból nyert adatok aktuálisak, de felvételük időigényes, a szekunder adatok nem biztos, hogy naprakészek, viszont gyorsan lehívhatóak.A primer és szekunder módszerek egyazon kutatásban ötvözhetőek, akárcsak a kvalitatív és kvantitatív módszerek, ezt nevezzük komplementeritásnak.

A 11.1.-es alfejezetben a kvalitatív (minőségi) módszereket, a 11.2.-es alfejezetben a kvantitatív (mennyiségi) módszereket ismertetem. A két módszer nem rangsorolható, mindkettőnek vannak előnyei és hátrányai.

Az előnyöket a 2. sz. Táblázatban foglaltam össze. Az adott módszerre eső választásunkat a kutatás céljának kell alárendelnünk, a választásunkat meg kell indokolni.

2. sz. Táblázat – A kutatási módszerek előnyei

A kvalitatív módszerek előnyei	A kvantitatív módszerek előnyei
Egy konkrét eset vagy jelenség hátterének alapos feltárása szöveges formában	Általános tendenciák kimutatása számadatok alapján
A megállapítások és eredmények az adott esetre vagy jelenségre vonatkoznak	A megállapítások általánosíthatósága magasabb fokú, feltéve, ha a mintánk kellőképpen reprezentatív
Nem szükségesek statisztikai ismeretek	A statisztikai programokkal kigenerált táblázatok és ábrák gyorsan növelik a dolgozat terjedelmét
Értelmezése egyszerű és megenged némi szubjektivitást	Értelmezése jártasságot feltételez a statisztikában, az eredmények látványosabban mutathatók be

Amennyiben a terjedelmi korlátok megengedik vagy a témaválasztásunk, esetleg kutatási célkitűzésünk indokolná, akkor mindkét módszert alkalmazhatjuk, de általában a szakdolgozatokban erre nincs szükség, elégséges vagy kvalitatív, vagy kvantitatív módszert hanszálni.

A kutatási módszer tehát az adatok típusára vonatkozik, melyen belül többféle eszközből is válogathatunk. A kutatás eszközei az adatgyűjtés konkrét, gyakorlati módjára vonatkoznak, melyekről a következő két fejezetben (11.1. és 11.2.) bővebben is szó lesz.

11.1. Kvalitatív kutatás

A kvalitatív kutatás „a jelenségek minőségi megismerésére, vélemények, attitűdök, hiedelmek összegyűjtésére és elemzésére irányul" (Boncz, 2015, 25.); lehetővé teszi a társadalmi folyamatok, jelenségek és problémák mélyebb megértését. A reprezentativitás nem szempont, így eredményeink és következtetéseink általánosíthatósági foka alacsonyabb, mint a kvantitatív módszerek esetében. Ez persze nem csorbítja a kutatás értékét, mert olyan információkhoz juthatunk, amelyekre egy mennyiségi elemzés során nem derül fény.

Egy korábbi, saját kutatásomat hoznám fel példaként a minőségi elemzésben rejlő fent említett potenciál kapcsán: a kutatás első szakaszában kérdőívekkel mennyiségileg felmértem, hogy a szociális munka szakos hallgatók hogyan viszonyulnak az eutanáziához. Kiderült, hogy a számszerűsített többség nem ért vele egyet. Azonban, a kutatás minőségi részéből megtudtam, hogy azon diákok, akiknek közvetlen környezetében élt terminális stádiumú beteg, vagyis közvetlenül megtapasztalták a betegséggel járó szenvedést, kivétel nélkül támogatnák az eutanáziát.

Amint az előbbi példából is láthatjuk, a kvantitatív és kvalitatív módszerek együttes alkalmazása révén egy sokkal összetettebb erdedményhalmazt kapunk, mintha csak az egyik, vagy csak a másik módszert alkalmaznánk.

A kvalitatív kutatás során általában az alábbi eszközöket különböztetjük meg:

1. Dokumentumelemzés: ennek során feltárjuk a szakirodalom álláspontját, levéltári forrásokat, intézményi honlapokon fellelhető információkat, stb. A dokumentumelemzés bizonyos mértékben minden dolgozatnak szerves része, még akkor is, ha főként kvalitatív módszereket szándékozunk alkalmazni.

2. Összehasonlítás: a minőségi összehasonlítás is szinte minden dolgozatban fellelhető, ugyanis ha például a szakirodalom alapján pro és kontra érveket vizsgálunk, fogalmat próbálunk konceptualizálni (több szerző véleménye, álláspontja alapján), saját következtetéseinket mérjük másokéhoz, máris összehasonlítást végzünk.

3. Megfigyelés: a megfigyelés során leírhatjuk a célcsoport vagy egyén viselkedését egy adott helyzetben. Végezhetjük akár a csoportba beépülve is, amikor például együtt játszunk gyerekekkel (részvételen keresztüli megfigyelés), vagy akár külső szemlélőként.

4. Interjú: egy célirányos beszélgetés képezi az alapját, többféle típusa létezik. Lehet irányított (strukturált) vagy félig irányított (félig strukturált) egy előre eltekészített vázlat szerint, de adott esetben lehet akár spontán, kötetlen egy téma körül. A spontán interjúkat főként egy adott témában való tájékozódáshoz szoktuk alkalmazni, a társadalomtudományi kutatásokban gyakoribb a strukturált vagy félig strukturált típus alkalmazása, mivel így célzott válaszokat kapunk a kutatási kérdéseinkre.

Strukturált és félig strukturált interjúkhoz előzetes vázlatot készítünk, melyet vagy a módszertani résznél, vagy a szakdolgozat függelékében tüntetünk fel.

5. Mélyinterjú vagy kvalitatív interjú. „A kvalitatív interjú egyedülállóan érzékeny és erős módszer arra, hogy megismerjük az alanyok mindennapi világukból merített tapasztalatait és e mindennapi világ megélt jelentéseit" (Kvale, 2005, 79. o.) A mélyinterjú tehát egy strukturált beszélgetés, mely során az alany feltárja érzéseit, belső gondolatait egy adott téma kapcsán.

6. A szakértői interjú, mint ahogy a neve is utal a tartalmára egy szakmát jól ismerő személlyel folytatott szintén strukturált beszélgetés).

7. Az életútinterjú során egy alany élettörténetét ismerhetjük meg. Fényt derülhet arra, hogy az alany által megélt események és környezeti hatások hogyank és miben befolyásolták.

8. Fókuszcsoport vagy csoportbeszélgetés: tulajdonképpen egy adott témára fokuszáló csoportos beszélgetés. Létszám tekintetében általában 6–12 fő vesz részt benne. Az interjút felvevő személy moderátorként egy vázlat alapján irányítja a beszélgetést. A kérdéseket nem direkt a résztvevőknek teszi fel, hanem arra ösztönzi őket, hogy beszélgessenek egymással, tegyenek föl kölcsönös kérdéseket és véleményezzék egymás nézőpontjait (Barbour–Kitzinger, 1999.). A kutató csak mederben tartja a beszélgetést. A beszélgetést akár rögzíteni is lehet, feltéve ha a csoport mindegyik tagja beleegyezik, de számítani kell arra, hogy ezáltal az alanyok feszélyezve érezhetik magukat és nem nyílnak meg. Az interjú után egy jelentés készül, melyben a kutató rögzíti az elhangzottakat.

9. Kísérlet: kevésbé jellemző a társadalmi problémák feltárásában, mivel csak korlátozottan alkalmazható. Az eredmények rendszerint látványosan bemutathatóak. Például kísérlettel megvizsgálhatjuk egy közösségi könyvtárban való részvételi hajlandóságot. Másik példa szerint bizonyos ingerek milyen reakciókat váltanak ki a célcsoport tagjaiból. Főbb lépései: mérés és észlelés, majd kiértékelés.

10. Esettanulmány: az esettanulmány során közvetlen (pl. megfigyelés és interjúk) és közvetett (az esettel foglalkozó szakemberek véleménye) módon szer-

zett információk alapján részletesen leírunk egy konkrét esetet, mint például egy hátrányos helyzetű egyénét vagy családét, majd ezt elemezzük.

11. Esetpélda: a esettanulmánynál terjedelmileg sokkal tömörebben ismertetjük a tényállást, majd levonjuk a következtetést pár mondatban.

A dolgozat módszertani részében fejtsük ki és indokoljuk a kiválasztott módszert és eszközöket, vagyis miért pont az adott módszert/eszközöket választottuk a kutatás célkitűzése szempontjából. Továbbá mutassuk be a kutatásban résztvevő célcsoportokat és indokoljuk meg, hogy miért pont őket választottuk.

FIGYELEM! Visszatérő hiba, hogy a hallgatók keverik a célcsoport és a minta fogalmát. Célcsoportról kvalitatív kutatás esetén beszélünk, mely nem reprezentatív. Mintáról és mintavételről kvantitatív kutatás esetén beszélünk, mely lehet reprezentatív vagy nem, illetve valószínűségi vagy nem.[4]

Javasolt irodalom a minőségi kutatáshoz:
Fedor Anita, Huszti Éva (szerk.). (2016). *Kutatásmódszertani kézikönyv*. Debreceni Egyetemi Kiadó. Debrecen.
Fónai Mihály, Kerülő Judit, Takács Péter (szerk.). 2002. *Bevezetés az alkalmazott kutatásmódszertanba*. Pro Educatione Alapítvány. Nyíregyháza.
Heltai Erzsébet, Tarjányi József. (1999). *A szociológiai interjú készítése*. TÁRKI. Budapest

11.2. A kvantitatív kutatás

A kvantitatív kutatás során számadatokkal dolgozunk, melyeket kérdőíves felmérés során nyerünk. Ezen adatokat adatbázisba kell rendeznünk, majd statisztikai műveletekkel értelmezzük a fennálló összefüggéseket. Az általános statisztika a matematika egyik ágazata, a társadalomtudományi kutatásokhoz való adaptációját társadalomstatisztikának nevezzük. Mivel módszerei a társadalmi viszonyok mérésének sajátosságaihoz vannak igazítva, többek között előfordulhatnak mérési hibák, amelyek matematikailag kevésbé pontos eredményekhez vezethetnek (Mezei és Veres, 2001). A továbbiakban a statisztika fogalom alatt a társadalomstatisztikát értjük.

A mennyiségi (kvantitatív) kutatás során kiválasztjuk a használni kívánt statisztikai programot. Az ingyenesen letölthető és használható statisztikai programok (PSPP, R, stb.) is megfelelnek a célnak, de az IBM által fejlesztett SPSS nem véletlenül kerül több ezer dollárba. Könnyebben használható és többet is tud, feltéve, ha tudunk szerezni hozzáférést. Az elmúlt 10-15 évben a kvantitatív kutatások színvonalával szemben állított követelmények növekedtek, ami főként a statisztikai programok elterjedésének tudható be.

4 A valószínűségi mintavétel a reprezentatív mintavétel egyik formája lehet. Azonban a reprezentatív mintavétel többet jelent, mint pusztán a valószínűségi mintavételt.

A kvantitatív kutatás során a kérdőívekben szereplő kérdések alapján változókat (pl. életkor, családi állapot, nem, iskolázottság, lakóhely stb) alakítunk ki. Általában egy kérdés egy változónak fel meg, de a komplex kérdések alapján előfordulhat, hogy egy kérdésből kettő, vagy akár több változó is származtatható. Az elemzés lényege, hogy a változókat bemutatjuk, azaz leírjuk, majd egymás között összehasonlítjuk. Így az adott kérdőívhez tartozó adatbázis belső összefüggéseit tárhatjuk fel.

További lehetőség, hogy egyazon alanyokon két vagy több időpontban (longitudinális vizsgálatnak is nevezzük), vagy különböző csoportokon alkalmazzuk le (kísérleti- és kontrollcsoport vizsgálat) ugyanazt a kérdőívet. Különböző csoportokat képezhetnek például egy oktatást célzó felmérésben a diákok és tanárok, vagy két tanügyi intézmény diákjai. Ezáltal azt vizsgáljuk, hogy a csoportok válaszai egymáshoz viszonyítva mutatnak-e mennyiségi eltéréseket. Amennyiben a különböző időpontok vagy csoportok lekérdezése mellett döntünk, célszerű ugyanazokat a kérdéseket és azonos mérési szinteket[5] alkalmaznunk. Ennek oka a statisztikai feldolgozhatóságban rejlik, mivel így sokkal könnyebb a változókat összevetni, műveleteket végezni.

Gyakran tapasztaltam a hallgatók által szerkesztett kérdőíveknél, hogy a két kiszemelt lekérdezendő csoportnak más-más kérdéseket fogalmaznak meg, vagy csak nagyon kis mértékben egyeznek a kérdőívek. Ilyenkor próbálom arra ösztönözni a hallgatóimat, hogy vizsgálják felül a kérdőíveiket és egységesítsék a kérdéseket; inkább kevesebb kérdést, vagy általánosabban megfogalmazott kérdéseket szerkesszenek, amelyeket mindkét csoportnak fel lehet tenni. Az ilyen eljárás során sokkal több összefüggést kinyerhetünk, mintha csoportok között egymással össze nem vethető kérdéseket tennénk fel.

A következő alfejezetekben azon statisztikai műveletekről lesz szó, amelyek segítségével kiértékelhetjük a kvantitatív adatokat. Az adott módszerhez fűződő magyarázatok célja az eddigi ismereteink felelevenítése. Mivel célkitűzésem, hogy a lehető legegyszerűbben ismertessem és magyarázzam a fogalmakat és műveleteket, ajánlom a statisztikai szakkönyvek használatát is.

11.2.1. Mintavételi eljárások

Elsőként tisztázzuk az alapsokaság és a minta fogalmát. Az alapsokaságot olyan egyének összessége alkotja, akik a kutatás tárgyát képezik. Az alapsokaságot, vagy más néven populációt a kutató fogja meghatározni a saját megfigyelési szempontjai alapján. Olyan csoportokat, közösségeket, vagy bizonyos jellemzőkkel rendelkező személyeket foglal magába, akiket tanulmányozni szeretnénk. Amennyiben ismerjük az alapsokaság attribútumait és szerkezetét, a kutatás során nem szükséges mindenkit lekérdezni. Elegendő egy az alapsokaságéval

5 Mérési szintek: intervallum, vagy arányskála, ordinális (rangsor), nominális (kategorikus).

megegyező szerkezetű mintát kialakítani, ebben az esetben beszélünk a minta reprezentativitásáról. Az alapsokaságból történő minta kialakítását nevezzük mintavételi eljárásnak.

A reprezentatív minta gyakorlatilag egy a teljes vizsgálati populáció, vagyis az alapsokaság jellemzőivel megegyező arányos lekérdezést feltételez. Az alapsokaság lekérdezése túlságosan idő- és költségigényes, ezért mintát veszünk belőle. A mintában szereplő egyénektől szerzett információk alapján a teljes alapsokaságra, vizsgálati populációra érvényes következtetéseket vonhatunk le. A mintában szereplő egyéneket kutatási célszemélyeknek vagy alanyoknak is nevezzük.

Ha jó a mintavételezési koncepciónk és valóban reprezentatív a mintánk, a kutatás eredményei megismételhetőek lesznek. Alkalmazhatunk valószínűségi vagy nem valószínűségi mintavételezési technikákat az eljárás során.

A **valószínűségi mintavételezés** azt jelenti, hogy minden alany egyenlő eséllyel kerülhet bele a mintába. Az egyes alanyok mintába való kerülésének esélye kiszámítható és véletlenszerű (random). A valószínűségi mintavételezés során az eredményeink hibahatára kisebb lesz. Ha a minta szerkezetileg is leképezi az alapsokaságot, akkor teljesül a reprezentatívitás kritériuma.[6] A két fogalom, vagyis a valószínűségi mintavételezés és a reprezentativitás tehát nem pont ugyanaz. A valószínűségi reprezentatív minta előnye, hogy a kinyert eredmények magas fokú megbízhatósággal fognak rendelkezni, hátránya viszont az, hogy a reprezentativitást elérése sokszor nem egyszerű. Inkább.

Valószínűségi mintavételezés során az alábbi technikákat használhatjuk:
– egyszerű véletlen mintavétel,
– szisztematikus mintavétel,
– rétegezett mintavétel,
– csoportos mintavétel,
– többlépcsős mintavétel,
– kombinált eljárások.

A **nem valószínűségi mintavételezés** során az alanyok mintába való bekerülésének esélyeit nem tudjuk kiszámítani, a kutató szubjektív megítélésén múlik. Ezáltal a reprezentativitás kritériumai nem fognak teljesülni, valamint a hibahatárt sem lehet pontosan kiszámolni. Előnye, hogy időhatékony és kis ráfordítással kivitelezhető, valamint bizonyos esetekben csakis ez a technika használható (amikor nem ismerjük a teljes populáció szerkezetét vagy nagyon kisszámú az alapsokaság és nem lehetséges a teljes lekérdezés).

Nem valószínűségi mintavételezés során az alábbi technikákat használhatjuk:
– könnyen elérhető alanyok,

6 A minta reprezentativitása bizonyos vizsgálati szempontok alapján értendő. A gyakorlatban nagyon ritkán tudunk a populáció minden jellemzőjére tekintettel reprezentatív mintát kialakítani.

– szakértői mintavétel,
– rendszertelen mintavétel.
– kvótás mintavétel,
– estei (*modal instance*) mintavétel,
– sokféleségi mintavétel,
– hólabda módszer.

Tapasztalataim szerint a hallgatók legkedveltebb mintavételezési technikája a hólabda módszer, viszont általában hibásan alkalmazzák. Jellegzetes hiba, hogy a hallgató saját maga megy oda a „hólabda" módszer szerint kiválasztott összes alanyhoz, holott a módszer lényege abban áll, hogy nem a kutató, hanem az alany fog rámutatni a soron következő kutatásban résztvevő egyénre. További hiba, hogy összekeverik a valószínűségi mintavételezéssel és reprezentatívként tüntetik fel a mintát, holott valójában nem az.

A hallgatók számára az egyik legkézenfekvőbb mintavételezési technika a könnyen elérhető alanyok lekérdezése, ezt használjuk, amikor saját ismerőseinket kérdezzük meg bármely közösségi oldalon. Noha rövid idő alatt viszonylag sok személyt (esetenként több százat) is elérhetünk, a mintánk magas esetszáma még nem garantálja a reprezentativítást. Ennek a módszernek az a fő korlátja, hogy ismerőseink, barátaink hasonló körülmények között szocializálódtak, tapasztalataik hasonlóak, többségük hasonló korosztályból való, személyes preferenciáik és gondolkodásmódjuk is hasonló. Ennek ellenére a közösségi oldalakon végzett kutatás is lehet elfogadható, viszont a módszert helyesen kell besorolni, valamint az ebből következő hátrányokat meg kell említeni a kutatás korlátainál. A mintavételi technikát is ajánlott előre egyeztetni a témavezető tanárral.

Tisztában kell lennünk azzal is, hogy amennyiben a kutatás során online lekérdezést alkalmazunk, a reprezentativitás elve aligha teljesül. Online környezetben nem biztos, hogy a teljes populáció behatárolható, sőt, ha nyilvánosan elérhetővé tesszük a kérdőívet, nem lesz ráhatásunk arra, hogy ki tölti ki. Különösen problémás a kérdőív saját közösségi média profilunkon történő megosztása, mivel első körben a saját barátaink, korcsoportunk és hasonló érdeklődésű és végzettségű ismerőseink fogják kitölteni. Amennyiben a kérdőívet tovább osztják, a válaszadók diverzifikálódhatnak, de az eredmények ekkor sem lesznek általános érvényűek és értékesek sem.

Az online lekérdezés esetenként lehet akár reprezentatív is, mint például amikor egy újság a saját olvasóit szondázza meg vagy egy cég a saját ügyfeleinek véleményére kíváncsi, mivel itt a teljes olvasói- vagy ügyfél-populáció ismert és elérhető.

11.2.2. A kérdőív szerkesztése

A kvantitatív módszer leggyakrabban használt eszköze a kérdőív, illetve a kérdőíves felmérés. Lebonyolítási módját tekintve beszélhetünk:

– önkitöltős kérdőívről, amikor az alany önállóan felel a kérdésekre,

– kérdezőbiztos által felvett kérdőívről, amikor a kutató/kutatók lekérdezik az adatokat,

– levélben, telefonon, interneten vagy egyéb kommunikációs eszközökön kiküldött kérdőívekről.

A kérdőív megszerkesztésekor a következőkre érdemes hagyatkozni:

1. előzetes tájékozódás: előzetesen ismernünk kell a kutatási területen a szakirodalomban felvetett dilemmákat; ha nem olvastunk utána a témának, jól kérdezni sem fogunk tudni.

2. meglévő kutatási eredmények feltérképezése: keressünk és tanulmányozzunk át hasonló kutatásokban alkalmazott kérdőíveket, amelyekből inspirálódhatunk; ez a későbbiekben is hasznunkra fog válni, mert a saját eredményeinket lesz mivel összevetni.

3. figyeljünk a statisztikai feldolgozhatóság követelményeire: olyan formában tegyük fel a kérdéseinket és adjuk meg a válaszlehetőségeket, hogy könnyen lehessen azokat kódolni, majd feldolgozni (lásd lentebb a kérdőív megszerkesztésére vonatkozó javaslatokat).

4. ha elkészítettük a kérdőívet, mielőtt lealkalmaznánk, kérjük ki a vezető tanárunk véleményét, akinek feltehetően rutinja és rálátása van a kutatásra.

A kérdőív szerkesztését tekintve négy tartalmi részt különíthetünk el:

1. demográfiai jellemzők;
2. szociokulturális jellemzők;
3. szociogazdasági jellemzők;
4. a témánkra irányuló konkrét kérdések.

A kérdőív szerkesztésekor tisztázzuk, hogy az adott kérdés melyik tartalmi részbe sorolható. A fenti jellemzők a dolgozat strukturálásakor, az eredmények bemutatásának sorrendjében is lehetséges szempont lesz.

A témánkra irányuló konkrét kérdéseket a szakirodalom előzetes ismerete nyomán fogalmazzuk meg, így a várható eredményeink könnyen összehasonlíthatóak lesznek más, korábbi kutatások eredményeivel. Mindez hozzájárul a dolgozatunk egységéhez, vagyis az elméleti rész és a gyakorlati rész kapcsolódni fog egymáshoz.

A feltett kérdések száma befolyásolja a kitöltéshez szükséges időt, ezáltal az alanyok válaszadási hajlandóságát is. Egy államvizsga dolgozathoz elégséges akár egy 15-20 kérdésből álló kérdőív is, egy mélyebb és átfogóbb tanulmány-

nál, ahol több idő jut a kérdőívek felvételére, elmehetünk a 30-45 kérdésig, ami egyénileg még feldolgozható. Fontos, hogy csak a feltétlenül szükséges kérdéseket tegyük fel, megfelelő számú válaszlehetőséggel.

Az adott kérdésre kapott válaszok rendszerint egy-egy változónak (*variable*) számítanak, de előfordulhat az is, hogy az adott kérdésünk az adatok kódolása során több változóra fog bomolni. A változó tulajdonképpen szimbolikus jelölése egy mennyiségnek vagy kategóriának. Például, ha rákérdezünk a kutatási alany életkorára, akkor a kérdést egyetlen változóként kezeljük. Ha viszont rákérdezünk arra, hogy *„az elmúlt egy évben milyen gyakran vett részt kulturális tevékenységekben"*, felsorolva néhány lehetséges tevékenységet, akkor miden egyes tevékenység külön-külön változóként fog szerepeli. A változók típusaira visszatérünk az adatbevitellel és az adatbázis rendezésével foglalkozó fejezetekben.

A kérdőív megszerkesztésére vonatkozóan az alábbi javaslatokat tenném, annak érdekében, hogy statisztikailag könnyebb legyen a kiértékelés:

1. Kerüljük a nyílt kérdéseket – ezeket inkább minőségi kutatásnál használjuk pl. az interjúknál, mert nagymértékben megnehezítik az adatok (válaszok) kódolását. Ha mégis nyílt kérdéseket tennénk fel, gondoljuk el előre, hogy a lehetséges válaszokat hogyan tudnánk véges számú kategóriákba rendezni. Például az alanyok foglalkozása rendszerint egy olyan kérdés, amire számtalan válaszlehetőség érkezhet. Hagyatkozhatunk azonban a foglalkozások nemzetközi besorolására, mely standardjai szerint bármilyen foglalkozás a 10 fő csoportba kódolható (ISCO-88).

2. Kétértékű (dihotóm) változókat, mint például FIÚ/LÁNY, FALU/VÁROS, IGEN/NEM, csakis akkor használjunk, amikor feltételül szükséges.

3. Alkalmazzunk minél több skála és intervallum típusú változót, például 1-től 5-ig számozva.

4. A skáláink és intervallumaink iránya ugyanaz legyen, az 1-es a kicsi, jelentéktelen, nem jellemző válaszokat jelölje, az 5-ös a nagy, legjelentősebb, legjellemzőbb válaszokat.

Hasznos linkek:
A foglalkozások nemzetközi besorolása elérhető az International Labour Organization (ILO) honlapján: https://www.ilo.org/public/english/bureau/stat/isco/isco88/ (hozzáférés 2023.01.12-én)

11.2.3. Az adatbevitel

Amennyiben papír alapú kérdőíveink vannak, az első lépés az összes kérdőív sorszámozása. A sorba rendezett kérdőívek legelső példánya fog „mesterpéldányként" szolgálni az adatok bevezetésekor.

A „mesterpéldányon" elvégezzük a válaszok numerikus kódolását, vagyis minden válaszlehetőséget megszámozunk. A számozás iránya mindig ugyanaz legyen, az 1-es a jelentéktelen, nem jellemző válaszokat jelölje, az 5-ös a legjelentősebb, legjellemzőbb válaszokat. Lehetőleg a zérós értéket sehol ne használjunk, 1-estől kezdjük a válaszok kódolását. A mesterpéldányt előttünk tartva, hogy emlékeztessen a kódolásra, sorban bevezetjük az összes kérdőívet. Minden kérdőívben szereplő adat egy sorban fog megjelenni, minden oszlopba az adott kérdésre kapott válaszok numerikus kódolása kerül.

Figyeljük arra, hogy az adott sorszámú kérdőívünk a statisztikai programban is azonos sorszámmal szerepeljen, így a későbbiekben, a sorszámok alapján visszakeresve könnyedén ki tudjuk javítani az esetleges hibákat, elgépeléseket. Ezen hibák az adatbázis „rendezésekor" általában láthatóvá válnak.

Az adatbevitel leggyorsabb módja, ha online kérdőívet alkalmazunk olyan platformon, amely a válaszokat adatbázisba rendezi és letöltési lehetőséget is biztosít. Ilyen például a *Google forms*, ahol a Válaszok (*Responses*) fülre kattintva *.csv formátumban letölthetőek saját számítógépünkre. Ezt a formátumot (*comma separated values*) például az *Excel* táblázatkezelő is ismeri. A kapott válaszok szöveges formátumban (*string*) fognak megjelenni, ahogyan a kérdőívünkben is szerepeltek. A táblázatban minden válasznak saját oszlopot alakítunk ki, majd numerikus formába kódoljuk (*pl. a helyettesítés/replace funkcióval*). Az adatbázisunkkal akár az Excelben is végezhetőek bizonyos műveletek, viszont a számolásokat érdemesebb a továbbiakban statisztikai programmal végezni (SPSS, PSPP, R stb.). Az adatbázisunk tehát egy táblázat lesz, melyben a sorok és oszlopok metszéspontját *celláknak* nevezzük.

A numerikus adatbázisunkat *copy-paste* módszerrel átvihetjük az általunk használt statisztikai programba.

Minden műveletnél megadom az SPSS és a PSPP-ben is követendő parancssort. A jelen kézikönyvben a szemléltető ábrák többségeben, illetve a beállításokhoz és a kimeneti ablakok értelmezéséhez a PSPP programot használom, mivel a legtöbb hallgató számára ez az elérhető opció.

Mindkként, fent említett statisztikai programban (SPSS és PSPP) két fő ablak vagy nézet közül választhatunk: a változók nézete (*Variable view*), valamint az adatok nézete (*Data view*) – melyek között a képernyő bal alsó sarkában válthatunk. Az adatnézet (*Data view*) ablakban a változóink numerikus értékeit fogjuk bevinni, ahol a sorok az egyes esetek (alanyok) válaszait tartalmazzák, az oszlopok pedig az egyes változókat tartalmazzák. A változó nézet ablakban (*Variable*

view) a változóinkat fogjuk konfigurálni: megadjuk a változó nevét, címkéjét, típusát, lehetséges értékeit stb.

A változók konfigurálása azért fontos, mert segíti az adatbázisban való tájékozódásunkat, ezáltal a műveleteket könnyebben tudjuk áttekinteni.

A változók nézete (4. sz. Ábra) a legtöbb statisztikai programban, hasonlóan néz ki.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	változó1	Numeric	8	0	A változó cimkéje	None	None	8	Right	Scale	Input
2										Scale	
3										Ordinal	
4										Nominal	

4. sz. Ábra – Változók nézete a statisztikai programban

A beállításokról a következőket érdemes megjegyezni:

Sorszám: a változók száma, vagyis általában a kérdés sorszáma;

Name: a változó rövidített neve, ékezetek és szóközök nélkül; csupán a tájékozódást segíti a műveletek végzése során;

Type: ajánlott a numerikus típus használata, amennyiben sikerült számokkal kódolni a kérdésre adott lehetséges válaszokat;

Widht: a változó értékének maximális hossza, alapértelmezés szerint 8 számjegy;

Decimals: az alapértelmezett két tizedes pontosságot érdemes átállítani 0 tizedes pontosságra, (amennyiben az adataink megengedik), így az adatok nézete átláthatóbb lesz;

Label: a változó címkéje, ékezetekkel és szóközökkel; ez a címke jelenik majd meg a kigenerált táblázatokban;

Values: a beállítások talán legtöbb figyelmet igénylő része, ahol az adott kérdésre kapott válaszlehetőségeket kódoljuk (a lehetséges értékekhez címkéket társítunk, pl. 1 – fiú, 2 – lány, vagy 1 – soha, 2 – ritkán, 3 – időnként, 4 – gyakran, 5 – állandóan);

Missing values: a hiányzó adatok arra utalnak, hogy az adott kérdésre nem kaptunk választ. Ezt az SPSS például úgy is tudja kezelni, ha az adott cellát üresen hagyjuk; vagy a hiányzó adatokat jelölhetjük olyan értékekkel, amik nem köszönnek vissza a válaszokban (konvencionálisan a „9”, „99”, „999” stb. használata is lehetséges, sőt bizonyos programokban, mint a PSPP szükséges is!). A kapott eredmények a hiányzó adatok nélkül, az érvényes válaszok (*valid responses*) alapján értendőek;

Collumns: az adatok nézetben megjelenő oszlophossz;

Alignment: a tartalom megjelenítésének helye a cellán belül (bal, jobb, közép);

Measure: a változó mértéktípusa, lehet skála (*scale*), sorrendi (*ordinal*) vagy névleges (*nominal*);[7]

Role: általában bemeneti, tehát input típusú adatok.

7 A mérési szintekről lásd a 11.2.6. alfejezetet.

11.2.4. Minták és csoportok

A kutatás során dolgozhatunk egy vagy több mintával. Mindegyik mintánk egy-egy bizonyos kutatási szempontok szerint kialakított populációt fog leképezni. Értelemszerüen ha lesz egy kontrollcsoportunk is, ők egy második mintát fognak alkotni. Ha különböző kategóriájú alanyokat vonunk be, annyi mintánk lesz, ahány kategóriájú alanyunk van. Több kategóriájú mintánk lesz, ha például egy kutatásban lekérdezzük a diákokat, tanárokat és szülőket (3 minta), vagy az orvosokat és pácienseket (2 minta).

A mintán belül is kialakíthatunk csoportokat, pl. nemek, életkor-kategóriák vagy egyéb, tetszőleges szempontok alapján. Ebben az esetben az adatbázisunk egyik változója fogja tartalmazni a csoportok szerinti felosztást.

Amennyiben két különálló mintánk van, szükséges lesz a mintákban megegyező változók összehasonlítása. Ilyen helyez akkor áll fenn, amikor a két vagy több különböző csoportnak ugyanazt a kérdést tettük fel és a válaszlehetőségek is azonosak.

Az adatok bevitelekor célszerű külön-külön adatbázisba kódolni mindegyik mintánkat. A statisztikai feldolgozás során viszont össze kell vonnunk őket egyetlen adatbázisba. Ilyenkor a változó nézetben mindkét mintára vonatkozó változókat beállítjuk, majd az adatokat az egyes változók megfelelő oszlopaiba visszük be.

*[DataSet2] — PSPPIRE Data Editor

File Edit View Data Transform Analyze Graphs Utilities Windows Help

Variable	Name	Type	Width	Decimal	Label	Value Labels	Missing Values	Columns	Align	Measure	Role
1	tanar1	Numeric	8	2	Tanár 1. kérdés	None	None	8	Right	Scale	Input
2	tanar2	Numeric	8	2	Tanár 2. kérdés						Input
3	tanar3	Numeric	8	2	Tanár 3. kérdés	A tanárokat leképező mintánk összes változója					Input
4	tanar4	Numeric	8	2	Tanár 4. kérdés						Input
5	tanar5	Numeric	8	2	Tanár 5. kérdés	None	None	8	Right	Scale	Input
6	diak1	Numeric	8	2	Diák 1. kérdés	None	None	8	Right	Scale	Input
7	diak2	Numeric	8	2	Diák 2. kérdés						Input
8	diak3	Numeric	8	2	Diák 3. kérdés	A diákokat leképező mintánk összes változója					Input
9	diak4	Numeric	8	2	Diák 4. kérdés						Input
10	diak5	Numeric	8	2	Diák 5. kérdés	None	None	8	Right	Scale	Input
11	szulo1	Numeric	8	2	Szülő 1. kérdés	None	None	8	Right	Scale	Input
12	szulo2	Numeric	8	2	Szülő 2. kérdés						Input
13	szulo3	Numeric	8	2	Szülő 3. kérdés	A szülőket leképező mintánk összes változója					Input
14	szulo4	Numeric	8	2	Szülő 4. kérdés						Input
15	szulo5	Numeric	8	2	Szülő 5. kérdés	None	None	8	Right	Scale	Input

5. sz. Ábra – PSPP adatbázis

Vegyünk egy példát, amiben három mintánk van: tanárok, diákok és szülők. Ebben az esetben a következő munkamenetet javasolom:

1. Elkészítjük a három különálló adatbázist, így az adott minta statisztikai elemzése adatbázisonként lehetséges. Ennek praktikus oka a hibalehetőségek és a véletlen adatvesztések minimalizálása.

2. Összevonjuk a mintáinkat egy közös adatbázisba. A változó nézetben (*Variable view*) egymás után, mintánként beállítunk minden változót (lásd a keretes részeket, 5. sz. Ábra).

Az adatnézet (*Data view*) az alábbiak szerint alakul (6. sz. Ábra):

*3 mintas adatbazis.sav [DataSet2] — PSPPIRE Data Editor

File Edit View Data Transform Analyze Graphs Utilities Windows Help

1 : szulo3 1,00

Case	tanar1	tanar2	tanar3	tanar4	tanar5	diak1	diak2	diak3	diak4	diak5	szulo1	szulo2	szulo3	szulo4	szulo5
1	1,00	6,00	2,00	1,00	6,00	2,00	1,00	6,00	2,00	1,00	6,00	2,00	1,00	6,00	2,00
2	2,00	5,00	4,00	2,00	5,00	4,00	2,00	5,00	4,00	2,00	5,00	4,00	2,00	5,00	4,00
3	2,00	4,00	1,00	2,00	4,00	1,00	2,00	4,00	1,00	2,00	4,00	1,00	2,00	4,00	1,00
4	3,00	[illegible]	[illegible]	[illegible]	[illegible]	3,00	[illegible]	[illegible]	[illegible]	3,00	3,00	[illegible]	[illegible]	[illegible]	3,00
5	4,00	[illegible]	[illegible]	[illegible]	[illegible]	2,00	[illegible]	[illegible]	[illegible]	4,00	4,00	[illegible]	[illegible]	[illegible]	2,00
6	5,00	[illegible]	[illegible]	[illegible]	[illegible]	2,00	[illegible]	[illegible]	[illegible]	5,00	6,00	[illegible]	[illegible]	[illegible]	2,00
7	7,00	[illegible]	[illegible]	[illegible]	[illegible]	3,00	[illegible]	[illegible]	[illegible]	7,00	5,00	[illegible]	[illegible]	[illegible]	3,00
8	8,00	[illegible]	[illegible]	[illegible]	[illegible]	4,00	[illegible]	[illegible]	[illegible]	8,00	5,00	[illegible]	[illegible]	[illegible]	4,00
9	5,00	5,00	4,00	5,00	5,00	4,00	5,00	5,00	4,00	5,00	5,00	4,00	5,00	5,00	4,00
10	4,00	2,00	5,00	4,00	2,00	5,00	4,00	2,00	5,00	4,00	2,00	5,00	4,00	2,00	5,00
11	2,00	2,00	1,00	2,00	2,00	1,00	2,00	2,00	1,00	2,00	2,00	1,00	2,00	2,00	1,00
12	1,00	6,00	2,00	1,00	6,00	2,00	1,00	6,00	2,00	1,00	6,00	2,00	1,00	6,00	2,00
13	1,00	4,00	1,00	1,00	4,00	1,00	1,00	4,00	1,00	1,00	4,00	1,00	1,00	4,00	1,00
14															

A tanárokat leképező minta változóinak értékei

A diákokat leképező minta változóinak értékei

A szülőket leképező minta változóinak értékei

6. sz. Ábra – PSPP adatbázis

3. A három minta megegyező változói között elvégezzük a statisztikai műveleteket, figyelembe véve a változók típusait.

A fentieket összefoglalva: ha egy mintán belül több csoportunk van, akkor az adott változó értékeit az eseti számoknak megfelelő egyetlen oszlopba írjuk. Ha több mintánk van, akkor a mintánként egymásnak megfelelő változókat külön oszlopokba tesszük.

11.2.5. Az adatbázis rendezése

Amennyiben az adatbázist online kérdőívezés nyomán hoztuk létre, kicsi a valószínűsége annak, hogy hibákat vétünk az adatbevitel során, feltéve, ha megfelelő volt a kérdőívek kódolása.

A papír alapú adatbevitel több hibalehetőséget rejt magában, mivel egy monoton tevékenység, mely folyamatosan fokozott figyelmet igényel. Ellenőrizzük tehát, hogy az „adatok” nézetnél ne legyenek olyan értékek, amelyek a válaszok kódolásánál nem fordultak elő, valamint javítsuk ki a hibákat a kérdőívek viszszakeresésével és ellenőrzésével, ha adódtak. Jól jöhet, ha megszámoztuk őket és sorrendben vezettük be.

A hibák többsége, amelyeket még nem sikerült azonosítanunk, a számítások során valószínűleg ki fog tűnni, sőt a gyakoriság táblázatokban vagy a kereszttáblákban is azonnal szemet szúrnak.

Az adatbázis rendezésére tekintsünk úgy, mint egy folyamatra, mivel a számítások megkezdését követően kisebb „utómunkákra” lesz szükség. Tehát az adatbázisunk létrehozásával és a hibák kijavításával még nem fejeztük be az adatrendezést.

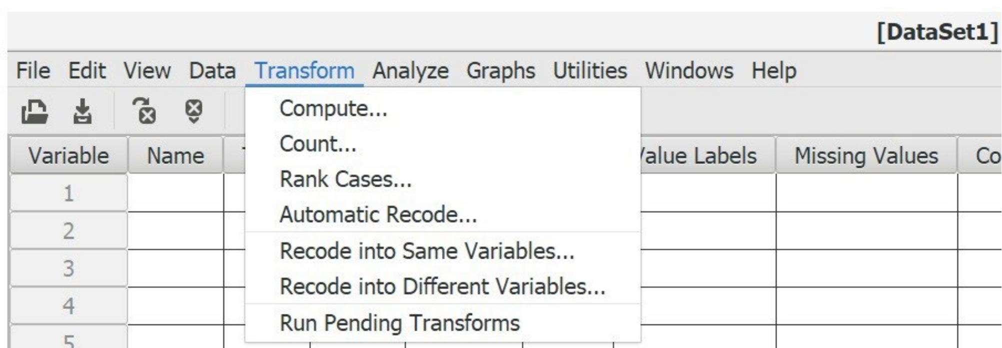
Az adatbázis rendezése során új változókat hozhatunk létre a meglévőkből, ezeket összevonhatjuk, vagy akár új értékekkel bíró kategóriákat is felállíthatunk. Erre a feldolgozási fázisra menet közben is bármikor visszatérhetünk, ha szükségét érezzük a változók átcsoportosításának, vagy újrakódolásának. Például olyankor lehet erre szükség, amikor egy gyakorisági táblázatban vagy egy kereszttáblában sok a zéró vagy a kis eseti számot megjelenítő cella.

A két leggyakrabban használt művelet a változók összevonása (*compute variable*) és a változók új változókba való újrakódolása (*recode into different variables*). Az azonos változóban történő újrakódolást (*recode into same variables*) lehetőleg mellőzzük, mert a meglévő adataink a művelet során véglegesen felülíródnak.

Változók összevonását például akkor használhatunk, amikor egy bizonyos jelenséget vagy jellemzőt több változón keresztül mértünk. Több változó pontszámának összege alapján új átlagpontszámot is számolni tudunk, adott esetben súlyozni egyes változókat. Például, ha az egyetemi oktatással való hallgatói elégedettségre vagyunk kíváncsiak, rákérdezhetünk, hogy a hallgatók mennyire elégedettek az átadott elméleti tudással, a szakmai gyakorlattal, az oktatói hozzáállással stb., így ez a néhány változó fogja leírni az általános elégedettséget. Ez is egy módja annak, hogy csökkentsük a kérdőívben szereplő kérdéseink számát.

A változók új változókba való újrakódolása során összevonhatunk bizonyos kategóriákat, vagy kiemelhetünk bizonyos válaszokat. Például, ha a kutatásban a dohányzási szokásokat mérjük fel, abból a skála típusú kérdésből, hogy *ki milyen gyakran dohányzik?*, visszafejthető, hogy kik a dohányzó és kik a nemdohányzó alanyok, ezáltal szintén „megspórolhatunk" egy kérdést. Másik példa, hogy a konkrét életkorokat (17, 26, 27, 21, 42, stb.) életkor-kategóriákba is rendezhetjük (25 év alatt, 25-35 között, 35-50 között és 50 felett).

A művelet elvégzése a PSPP és SPSS programokban hasonló módon történik. A 7. sz. Ábrán a PSPP felülete látható.



7. sz. Ábra – A PSPP statisztikai program felülete

Az SPSS egy kiforrottabb, több lehetőséget nyújtó program, viszont az alapműveletek elvégzésére a PSPP is tökéletesen alkalmas, ráadásul ingyenes. Ha a

fenti művelet bármelyike gondot okozna, vagy nem a hallgató szándéka szerint sikerülne, segítséget nyújthatnak pl. a Youtube-on elérhető *SPSS tutorial* videók. Ezt a lehetőséget a későbbiekben bármilyen művelet elvégzése kapcsán ajánlott igénybe venni.

Az adatbázis rendezése során ajánlott biztonsági másolatok készítése minden nagyobb tervezett átalakítás előtt, mivel – a tapasztalat azt mutatja – nem ritka eset, hogy a hallgatók véletlen adatvesztéssel, elcsúszó oszlopokkal, sorokkal szembesültek. A nem kívánt „adatbázis-balesetek" okozta károkat célszerű minimalizálni, mert sokkal könnyebb egy korábbi elmentett verzióra visszatérni, mint a kérdőíveket visszakeresni.

11.2.6. A változók fogalma, típusai

Az adatbázis rendezésének egyik feltétele, hogy tisztában legyünk a változók fogalmával. A változó gyakorlatilag egy bizonyos mérési szinten számszerűsített adat, a vizsgálati egység attribútuma. A változókat többféle szempont alapján is osztályozhatjuk.

1. A változók mérési szintje szerinti besorolása:

a. intervallum vagy arányskála (a statisztikai programban használt megnevezése *scale*): besorol egy változót a jellemzőjének mértéke alapján; a vizsgálati egységek egymástól azonos távolságra vannak (pl. hőmérsékleti fokok, életkor, időtartam).

b. ordinális (a statisztikai programban használt megnevezése *ordinal*): Két típusú ordinális változót különböztethetünk meg. Az *ordinális sorrendi* változók relatív attitűdöket, preferenciákat, véleményeket, percepciókat jelölnek; a válaszok mennyiségi alapon, nagyság szerint sorba rendezhetőek, de az eltérés mértéke szubjektív (pl. kicsi-nagy; olimpiai érmék: arany, ezüst, bronz). Tehát az elemek sorba rendezhetőek, de a köztük lévő arány vagy különbség egzakt módon nem értelmezhető (pl. a végzettség esetén). Az *ordinális rangsor* típusú változóknál az elemek rangjai szintén sorrendbe állíthatóak, de az elemek közti különbségek relatív arányai nem fontosak, nem relevánsak (pl. 1-től 5-ig terjedő skála). Ez utóbbi változótípus abban különbözik az arányskálától, hogy míg az arányskála esetében a különbségek és arányok pontosan meghatározhatók és relevánsak, addig az ordinális rangsor esetében csak a sorrendi információ releváns.

c. nominális, másképpen névleges vagy kategorikus (a statisztikai programban használt megnevezése *nominal*): diszkrét elemeket tartalmazó változó, olyan azonosító címkék alkotják, amelyek nem rangsorolhatóak, mivel önálló kategóriákat vagy csoportokat jelölnek (pl. igen/nem, fiú/lány, vallások, falu/város, foglalkozások, színek stb.).

2. Folytonos és diszkrét változók

Folytonos változók: a változó egy adott tartományon belül elvileg bármilyen végtelen számú értéket felvehet, amelyek sorrendbe állíthatóak. Az egyes értékek közti távolság is a végtelenségig bontható. A gyakorlatban a folytonos változók a skála típusú mérési szintre utalnak. Bizonyos esetekben az ordinális típusú változóinkat is folytonosnak tekinthetjük, pl. amikor öt vagy ötnél több rangsorolható kategóriát tartalmaznak. Példák lehetnek folytonos változókra: életkor, testmagasság, jövedelem stb. Grafikusan összekapcsolt pontokkal ábrázolható.

Diszkrét változók: a változó elemeinek száma véges, jól elhatárolható csoportokat, kategóriákat alakíthatunk ki, amelyek azonban nem rendezhetőek sorba (pl. nemek, családi állapot, testvérek száma, szinek). Grafikusan elszigetelt pontokkal ábrázolható.

3. Független és függő változók

Független változók: a kutatási alanyok állandó, adott tulajdonságai, melyeket a kutató előzetesen alakít ki (pl. csoport szerinti eloszlások, nemek, vallás stb.). A független változó képes befolyásolni a többi változó viselkedését, viszont ez fordítva nem áll fenn. A független változó az okot jelöli.

Függő változók: a kutatási alanyok magatartásában változtatható tényezők (pl. egy szokás gyakorisága). A függő változó a hatást jelöli.

4. Dihotóm és többértékű változók

Dihotóm változók: bináris változóknak is nevezzük, kettős értékű változók, azaz adott intervallumban csak két értéket vehetnek fel (pl. igen/nem);

Többértékű változók: tetszőleges számú értéket felvehetnek (pl., ha 5 fokozatú vagy intervallum skálákkal dolgozunk a kérdőívben, akkor 5 értékünk lehet).

A változók osztályozásának gyakorlati jelentősége abban rejlik, hogy egy adott műveletet rendszerint csak bizonyos típusú változókkal végezhetünk el.

11.2.7. Statisztikai alapfogalmak

A statisztikában is használt alapvető fogalmakat és egyezményes rövidítéseiket az alábbiakban ábécé sorrendben ismertetem:

Átlag - mean (*m*): középérték mutató, leggyakrabban számtani átlagot jelöl, a statisztikai sokaság tömörítésére szolgál.

Esetszám - number (*n*): az esetek számát jelöli.

Korrelációs együttható – correlation (*r*): két változó közötti lineáris kapcsolat nagyságát és irányát jelzi. Értékei (+1) és (-1) között alakulnak, ahol (+1) a

tökéletes együttmozgást, (0) a bárminemű összefüggés hiányát, (-1) a tökéletes ellentétes mozgást jelöli.

A korrelációs értékek 1–0.5-ig erősnek, 0.5–0.3 között közepesek, 0.3–0.1 között gyengének számítanak.

Standard eltérés – standard deviation (*SD*): általában az átlag mintavételi szórását jelenti, a szórás négyzetgyökével egyenlő. A standard hiba megmutatja, hogy az adatokból nyert átlag mennyire pontos.

Szignifiancia szint - significance level (*p*): a szignifikancia szintet a statisztikai hipotézisvizsgálathoz használjuk. A nullhipotézis egy olyan alapvető feltételezés, miszerint semmi sem történt, vagy változott a megfigyelt eredmény tekintetében. Az általánosan elfogadott α szint 5%, vagyis az ezen érték alatti „p" értékek ($p \leq 0.05$) a nullhipotézis elutasítását indokolják. Más szavakkal a nullhipotézis elutasítása azt jelenti, hogy a megfigyelt változás 95%-ban nem a véletlennek (vagy akár mérési hibának) köszönhető.

Szabadsági fokok - degrees of freedom (*df*): az egymástól függetlenül választható elemek számával egyenlő, értéke általában *n-1* (ahol *n* az eloszlásban szereplő elemek száma). A szabadsági fokok megadják, hogy hány adatpontból származik az adott számítás vagy próba.

Variancia - variance: más néven szórásnégyzet, megmutatja, hogy egy valószínűségi változó milyen mértékben szóródik a középérték körül. A valószínűségi változókkal lehet modellezni a véletlen jelenségeket, valószínűségi becslések készítésére használjuk őket. A variancia gyakorlatilag a változónk értékeinek eloszlását jelzi, esetenként (pl. t-teszt) vizsgálhatjuk, hogy két változó értékeinek eloszlása azonos-e vagy sem.

11.2.8. Statisztikai műveletek javasolt sorrendje

Miután kialakítottuk az adatbázisunkat és azonosítottuk a változók típusait, kezdődhet a számolás. A műveletek elvégzésekor az alábbi sorrendet javasolnám:

1. Kezdjük a leíró jellegű gyakoriság táblázatokkal (pl. bemutatjuk a populáció legfőbb jellemzőit);

2. Kereszttáblákat generálhatunk, melyeken megfigyelhetünk bizonyos összefüggéseket és tendenciákat – ez kiinduló alap lehet korrelációs összefüggések és szignifikancia szintek vizsgálatához;

3. Csoportokat állítunk fel (pl. korcsoportok szerint, nemek szerint, jellemzők szerint stb.) és vizsgáljuk a köztük levő összefüggéseket;

4. Közepes és erős korrelációs viszonyokat keresünk:

a. a demográfiai változóink és a konkrét kutatási kérdésink között,

b. a szocio-kulturális változóink és a konkrét kutatási kérdésink között,

c. a szocio-gazdasági változóink és a konkrét kutatási kérdésink között,

d. vizsgálhatjuk a fenti változókat akár egymás között is,

e. vizsgálhatjuk a konkrét kutatási kérdéseink közötti összefüggéseket;

5. Megnézzük, hogy bizonyos változóink összevonhatóak-e, vagyis leírhatóak-e egyetlen kompozit változóval (főkomponens és faktor analízis);

6. Vizsgáljuk, hogy bizonyos változóink képesek-e előrejelezni más változókat, vagyis prediktív modelleket állítunk fel (regresszió analízis).

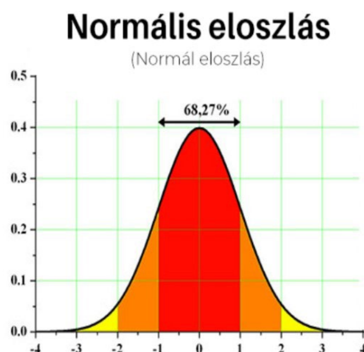
A fenti javaslattól eltérően más sorrend is elképzelhető, sőt gyakran előfordul, hogy az eredmények bemutatásakor, vagy a dolgozat strukturálásakor vesszük észre, hogy bizonyos összefüggéseket nem vizsgáltunk, bár lett volna rá adat. Ilyenkor visszatérhetünk az előző lépésekre és pótoljuk a hiányzó számításokat.

Ajánlott irodalom:
Pallant, July (2007), *SPSS Survival Manual*, Open University Press, Berkshire
Székely Mária, Barna Ildikó, (2002), *Túlélőkészlet az SPSS-hez*, Typotex Kiadó, Budapest

11.2.9. Az adatok eloszlásának vizsgálata

Mielőtt elkezdenénk a statisztikai számításokat, vizsgáljuk meg az adataink eloszlását. Ez a legelső művelet, amelyet azért kell elvégeznünk, hogy megtudjuk parametrikus vagy nem-parametrikus teszteket használhatunk a továbbiakban. A parametrikus és nem parametrikus tesztek a populáció eloszlásának megközelítési módjában különböznek, bővebben a 12.1. alfejezetben tárgyaljuk őket.

Minden egyes változónk az érvényes válaszok számával megegyező elemet tartalmaz. Ezek az elemek számértékeket vesznek fel (pl. 1-től 5-ig terjedő skála esetében). Ha az elemeket grafikusan egy görbén ábrázolnánk, és a legtöbb elem középre tömörülne, valamint sűrűségük szimmetrikusan csökkenne egy harangot ábrázolva (Gauss görbe), akkor az adatok normális eloszlásúak (8. sz. Ábra).



8. sz. Ábra – A normális eloszlás

Forrás: SPSSABC.HU (é.n.), letöltve 2022.01.12-én

Bár ez a fajta eloszlás a természetben nagyon gyakori, a kis mintás kutatásokban kifejezetten ritkának számít.

Az adatelosztást minden intervallum és arányskála esetében ellenőriznünk kell, hogy eldönthessük, milyen típusú teszteket végezhetünk az adott változókkal. A kategorikus mérési szintű változók esetén nem vizsgáljuk az eloszlást.

Normális eloszlás esetén parametrikus teszteket végezhetünk. Kategorikus változók nem mérhetőek ezekkel a típusú tesztekkel. Amennyiben az adataink eloszlása nem felel meg a normalitás kritériumainak (nem-normális eloszlásúak), akkor non-parametrikus teszteket végezhetünk. Továbbá normális eloszlású adatoknál nem számít, hogy átlaggal vagy mediánnal számolunk, a nem-normális eloszlású adatoknál a medián a jobb becslés a középértékre vonatkozóan.[8]

A normalitás vizsgálata Shapiro-Wilk teszttel és Kolmogorov-Smirnov teszttel történik. A teszt null-hipotézise, hogy az adatok eloszlása nem tér el a normális eloszlástól. A Kolmogorov- Smirnov (KS) és Shapiro-Wilk (SW) tesztek azt vizsgálják, hogy az adataink eloszlása illeszkedik-e a normális eloszlásra. Ha a teszt eredménye **nem szignifikáns** (p>.05), akkor normális eloszlású a változónk adatsora. Mindkét teszt használható, de a Shapiro-Wilk pontosabb eredményt ad és kis mintákhoz is megfelelő (n<50). A Kolmogorov-Smirnov teszt nagyobb minták esetén alkalmazható n>50.

Ha a ferdeség és csúcsosság értéke nullához közelít, akkor valószínű, hogy normális eloszlás jellemzi az adatainkat. Az említett tesztek hozhatnak nem szignifikéns eredményt, ha ferdeség és csúcsosság értéke nem nulla, viszont lényegesen és feltűnően nem tér el a normális eloszlástól.

Az SPSS felkínálja a normalitás vizsgálatának több lehetőségét is:

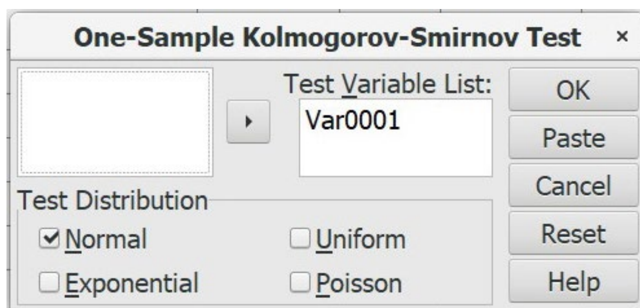
Első módszer:
Analyze → Descriptive Statistics → Explore
A *Dependent List* ablakba betesszük a vizsgálni kívánt változónkat, a *Factor List*-et üresen hagyjuk. A *Plots* fül alatt bejelöljük a *Normality plots with test* opciót.

Így kiszámolhatjuk a *Kolmogorov-Smirnov* és a *Shapiro-Wilk* tesztek eredményeit.

Második módszer:
Analyze→ Nonparametric Tests → Legacy Dialog → 2 Independent Samples
Ha két csoportunk van, akkor a *Test Variable List* ablakba betesszük a vizsgálni kívánt függő változónkat, a *Grouping Variable* ablakba betesszük a függő változónkat. A *Test Type* beállításoknál bejelöljük a *Kolmogorov-Smirnovn Z* probát.

8 Középértékek: átlag, módusz, medián.

PSPP-ben is elvégezhető a Kolmogorov-Smirnov teszt:
Analyze → Non parametric statistics → 1 Sample KS



A kiszemelt változónkat betesszük a *Test-Variable List* ablakba, kipipáljuk a *Test Distribution* opciók közül a *Normal*-t, majd lefuttatjuk a tesztet.

Az alábbi kimeneti ablak jelenik meg a képernyőn:

NPAR TESTS

NPAR TEST
/KOLMOGOROV-SMIRNOV (NORMAL) = Var0001.

One-Sample Kolmogorov-Smirnov Test

		Var0001
N		20
Normal Parameters	*Mean*	4,20
	Std. Deviation	2,76
Most Extreme Differences	*Absolute*	,17
	Positive	,17
	Negative	-,12
Kolmogorov-Smirnov Z		,75
Asymp. Sig. (2-tailed)		,626

A keretezett részt értelmezve: mivel a teszt szignifikancia szintje nagyobb, mint .05, vagyis nem szignifikáns, megállapíthatjuk, hogy a változónk nem tér el a normális eloszlástól. Ebből következik, hogy az adott változóval parametrikus teszteket végezhetünk.

12. STATISZTIKAI MŰVELETEK

12.1. Paraméteres és nem-paraméteres tesztek

Miután megvizsgáltuk az összes intervallum és arányskála típusú változónk adateloszlását, következhet a paraméteres és nem paraméteres technikák kiválasztása.

A **paraméteres tesztek** jellemzője, hogy feltételezi a vizsgált változó ismert normális eloszlását. Az ebbe a csoportba sorolható próbák csoport átlagokon alapulnak. A paraméteres teszteknek nagyobb a statisztikai erejük, mint a nem-paraméteres teszteknek. Folytonos mérési szintű változók esetén használhatjuk.[9] Kategorikus (nominális) mérési szintű változók esetén nem használhatjuk. Az ordinális (rangsor) típusú változók okozhatnak némi fejtörést, de általában szintén nem alkalmasak a paraméteres próbákhoz. Mivel kivételes esetekben kezelhetjük őket folytonos változóként is[10], ilyenkor feltételesen használhatunk paraméteres tesztet is, feltéve, ha nem helyettesíthetőek nem-paraméteres próbával.

A **nem paraméteres tesztek** jellemzője, hogy az adatsor nem-normális eloszlású. Az ebbe a csoportba sorolható próbák mediánokon alapszanak. A nem paraméteres tesztek általában úgy működnek, hogy sorrendbe állítják az összes értéket (rangsorolnak), majd a rangokkal[11] számolnak.

Kategorikus (nominális) és ordinális (rangsor) mérési szintű változók esetén használhatók.

Ha normális eloszlású adatokkal nem parametrikus teszteket végzünk, gyakorlatilag a parametrikus teszttel megegyező eredményt kapunk, de ez fordítva már nem áll fenn.

A statisztikai próbák kiválasztásakor három kérdés megválaszolása lehet a segítségünkre:

1. Milyen összefüggéseket szeretnénk vizsgálni?
2. A vizsgálat tárgyát képező változóink milyen típusúak?
3. Az összefüggés vizsgálatára milyen statisztikai művelet lehet alkalmas?

---

9 Nem normál eloszlások esetén is megbízható eredményeket tud nyújtani, feltéve, ha a minta nagysága megfelelő. Frost szerint 1 mintás t-teszt esetén minimum 20 megfigyelés, 2 mintás t-tesz esetén csoportonként minimum 20 megfigyelés, míg egyutas ANOVA esetén 2-9 csoportig minimum 15 megfigyelés és 10-12 csoportig minimum 20 megfigyelés szükséges.

10 Ha az ordinális kategóriák nem esnek túl messze egymástól (pl. iskolai osztályok vagy évfolyamok) és legalább öt vagy ötnél több elemet tartalmaznak, akkor tekinthetjük őket folytonos változónak.

11 A rangok gyakorlatilag sorszámok.

1. Milyen összefüggéseket szeretnénk vizsgálni?

Mivel minden kutatás egyedi, ezért maga a kutató dönti el, hogy mely változók közötti összefüggéseket szeretné feltérképezni. Erre általános módszer nincs, viszont hagyatkozhatunk a logikánkra. Meg kell vizsgálnunk, hogy a két (vagy akár több) változónk közötti összefüggésnek van-e, vagy lehet-e egyáltalán értelme. Például az alanyok jövedelme és végzettsége között lehet értelmes összefüggés, de az alanyok jövedelme és bőrszíne között biztosan nincs. Bár elvileg ez utóbbi is számolható, viszont az eredménynek nem lesz értelme. (További támpontot nyújthat a 11.2.8. alfejezetben javasolt vizsgálati sorrend.)

2. A vizsgálat tárgyát képező változóink milyen típusúak?

Mivel a változókat többféle szempont alapján lehet besorolni, ezért meg kell vizsgálni, hogy az elvégzendő statisztikai próba milyen típusú változókat feltételez. Elsőként figyelnünk kell az adatok eloszlására, innen tudjuk meg, hogy melyik – a parametrikus vagy a nem parametrikus módszer – közül fogunk választani.

A továbbiakban eldöntendő, hogy a változónk folytonos vagy kategorikus[12], illetve függő vagy független.

A folytonos változókról már tudjuk, hogy általában egy skáláról vagy intervallumról van szó. Az ordinális (sorrendi vagy rangsor) típusú változóink helyes besorolása okozhat némi gondot, ugyanis esetenként kezelhetjük őket úgy is, mint folytonos változókat.[13]

A kategorikus változók beazonosítása egyszerű, de még különbséget kell tennünk – a statisztikai próbák kiválasztásának szempontjából – a dihotóm (más szóval bináris vagy két értékű) és a több értéket felvevő változók között.

Bizonyos teszteknél szükségünk lehet arra is, hogy eldöntsük, az adott változó-pár közül melyik a független és melyik a függő változónk, itt célszerű a minta állandó jellemzőit független változónak tekinteni (pl. demográfiai jellemzők), a kutatás témájára vonatkozó kérdésekre adott válaszokat függő változónak.[14]

3. Az összefüggés vizsgálatára milyen statisztikai művelet lehet alkalmas?

Ez a lépés viszonylag egyszerű, mivel a lehetséges műveletek feltételei adottak. A 3. sz. Táblázat segítségünkre lehet a megfelelő próba kiválasztásában. A táblázatba foglalt statisztikai próbákkal változók közötti kapcsolatot, csoportokon belüli és csoportok közötti összefüggéseket vizsgálhatunk. Tájékozódási támpontunk lehet a mérések és csoportok (populációk) száma, illetve a változók típusa.

12 Lásd a 11.2.6 alfejezetet.

13 Ezt a kutatónak kell eldöntenie, de támpont lehet, ha az ordinális kategóriák nem esnek túl messze egymástól (pl. iskolai osztályok vagy évfolyamok) és legalább öt vagy ötnél több elemet tartalmaznak. Ha viszont különböző képzési szinteket jelölnek (általános iskola, középiskola, líceum, egyetem, posztgraduális képzés), akkor célszerűbb kategorikus változóként kezelni.

14 A függő-független változó dilemmája esetfüggő is lehet, tehát itt is a logikára alapozunk.

3. sz. Táblázat – Gyakori statisztikai próbák paraméteres és nem paraméteres megfelelőkkel

Mit mérünk?	**Paraméteres próbák**	**Nem paraméteres próbák**
Kategorikus változó értékeinek eloszlása (arányulása) annak specifikus részeihez, vagy az elvárt értékekhez képest.	–	Khi-négyzet illeszkedésvizsgálat (Chi-square for goodness of fit)
Kategorikus változók egyes elemei közti gyakoriságok szignifikánsan különböznek-e vagy sem.	–	Khi-négyzet függetlenségpróba (Chi-square for independence) Vagy Fisher féle egzakt teszt (Fisher's Exact Test)
Azonos populáción belül két kategorikus bináris változónak különbözik-e az eloszlása? Ismérve a 2x2-es kontingencia mátrix. (pl. két különböző időpont)	–	McNemar próba (McNemar' Test)
Azonos populáción belül legalább három kategorikus bináris változónak különbözik-e az eloszlása? (pl. 3 különböző időpont)	–	Cochran féle Q próba (Cochran's Q Test)
Két értékelő közötti nézetkülönbséget vagy nézetazonosságot vizsgálja. Szükséges két kategorikus változó, azonos számú kategóriával.	–	Cohen féle kappa együttható (Cohen's Kappa Measure of Agreement)
Két független csoport közötti különbségek vizsgálata egy folytonos változóhoz viszonyítva.	Független mintás t-teszt (Independent-samples t-test)	Mann-Whitney U próba (Mann-Whitney U Test)
Azonos populáción belül az alanyokat két különböző időpontban vagy körülmény mellett mérjük fel egy folytonos változóhoz viszonyítva.	Páros mintás t-teszt (Paired-samples t-test)	Wilcoxon féle előjeles rang teszt (Wilcoxon Signed Rank Test)
Három vagy több független csoport közötti különbségek vizsgálata folytonos változóhoz viszonyítva.	Egyirányú (egytényezős) csoportok közti ANOVA (One-way between-groups ANOVA)	Kruskal-Wallis próba (Kruskal-Wallis Test)
Azonos populáción belül legalább három folytonos vagy ordinális változó eloszlását vizsgálja.	Egyiányú (egytényezős) egymintás ANOVA (One-way repeated-measures ANOVA)	Friedman próba (Friedman Test)

A fenti táblázat a társadalomtudományokban gyakran alkalmazott próbákat gyűjtötte össze, a teljesség igénye nélkül. A kutatás igényei szerint alkalmazhatunk más próbákat is.

12.2. Leíró és következtető statisztikai technikák

A kvantitatív adatokat statisztikai elemzéssel dolgozzuk fel. Az elemzés két részre bontható: a leíró statisztika és a következtető statisztika.

A leíró (*descriptive*) statisztika az adott minta elemzéséhez nyújt segítséget, a következtető (*inductive*) statisztika a teljes populáció jellemzőit próbálja valószínűség számítási alapon kikövetkeztetni.

A kutatás eredményeinek bemutatása során részletesen megvizsgáljuk és leírjuk a mintát, mint például az alanyok demográfiai és egyéb szociokulturális és szociogazdasági jellemzőit. Amennyiben kettő vagy több csoportunk, esetleg kontrollcsoportunk is van, akkor mindegyik csoport külön-külön mintát képez, ezért mindegyik mintánkat részletesen leírjuk.

A teljes populációra vonatkozó következtetések becsléseken és hipotézisek vizsgálatán alapulnak, ezeket a következtető vagy más néven induktív statisztikai módszerekkel tudjuk kiszámítani. A gyakorlatban a következtető statisztikai számítások során is alkalmazhatunk leíró módszereket, tehát a két módszer elkülönítése inkább elméleti jellegű.[15]

12.3. Egy és többváltozós elemzés

A leíró statisztika egyváltozós elemzése elhelyezkedési és szóródási mutatók vizsgálatán alapszik.

Elhelyezkedési mutatók:

a. átlag: az értékek számtani középarányosa, a számsor értékeit összeadjuk, majd elosztjuk az elemszámmal;

b. medián: számtani középérték, melyet úgy kapunk, hogy az elemeket növekvő sorrendbe rendezzük és a sorozat közepén lévő elemünk lesz a medián;

c. módusz: egy számtani sorozat esetében a módusz a leggyakrabban előforduló elem;

d. kvantilis: olyan pontok, amelyek a mintát bizonyos arányban kétfelé osztják, mint például felező (medián), negyedelő (kvartilis) és századoló (percentilis) pontok.

Szóródási mutatók:

a. szórás: a középértéktől való eltérést mutatja;

b. variancia: szórásnégyezet;

c. terjedelem: a legnagyobb és legkisebb elem közti különbség.

Eloszlás alakjának mutatói:

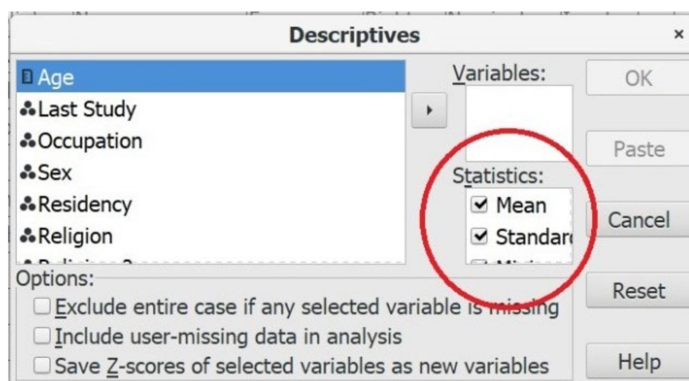
a. ferdeség (*skewness*): a változó elemeinek eloszlásának a szimmetriájára utal a normális eloszláshoz képest (a normális eloszlás görbéje szimmetrikus alakzatú);

15 Az eredmények bemutatásakor nem szükséges megjelölni, vagy elkülöníteni, hogy leíró vagy következtető statisztikai módszerrel jutottunk az adott eredményre, összefüggésre.

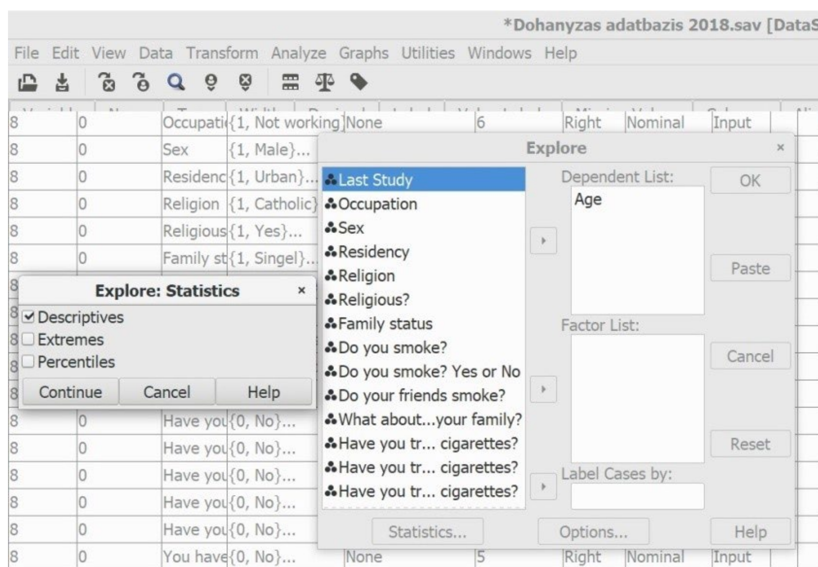
b. csúcsosság (*kurtosis*): a változó elemeinek eloszlásának alakzata mennyire lapos vagy csúcsos a normális eloszláshoz képest.

Egyváltozós elemzést többféleképpen végezhetünk a statisztikai programokban, az alábbiakban három lehetőséget javasolunk. A PSPP és az SPSS parancssora megegyezik, a beállításoknál van némi különbség.

Analyze → Descriptive Statistics → Frequencies[16] (itt beállítjuk azokat a mutatókat, amelyeket szeretnénk kiszámolni), vagy
Analyze → Descriptive Statistics → Descriptives



Harmadik lehetőség, ahol több információt jeleníthetünk meg:
Analyze → Descriptive Statistics → Explore



A *Statisztikák* fül alatt kiválasztjuk a *Descriptives* opciót, majd a *Continue* és *OK* gombokra kattintunk. A következő eredményeket kapjuk a kimeneti ablakban:

16 A gyakoriság számításánál fogjuk részletezni a beállítási lehetőségeket a 12.4.1. alfejezetben.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Age	312	99,0476%	3	0,952381%	315	100%

Descriptives

			Statistic	Std. Error
Age	Mean		24,35	,34
	95% Confidence Interval for Mean	Lower Bound	23,69	
		Upper Bound	25,01	
	5% Trimmed Mean		23,52	
	Median		23,00	
	Variance		35,49	
	Std. Deviation		5,96	
	Minimum		18,00	
	Maximum		55,00	
	Range		37,00	
	Interquartile Range		5,00	
	Skewness		2,61	,14
	Kurtosis		8,27	,28

A kapott eredmények közül csak azokat emeljük át a dolgozatunkban, amelyekre vizsgálati szempontunkból szükségünk van.

A többváltozós elemzés során az egynél több változó közötti kapcsolatot vizsgáljuk. Az elemzés során leíró és következtető módszereket is alkalmazunk.

12.4. Leíró statisztikai technikák

A leíró statisztikai számítások során minden változót külön-külön megvizsgálunk, azaz „leírjuk” annak jellemzőit. Ezáltal a mintánk áttekinthetőbb lesz, sőt lehetőségünk nyílik az adatok vizualizációjára is, különböző grafikonok és diagramok segítségével. A leíró statisztika egyváltozós és többváltozós elemzést is lehetővé tesz. Elhelyezkedési (pl. átlag, medián, módusz, percentilis és kvantilis) és szóródási (pl. szórás, variancia, terjedelem) mutatók jellemzik.

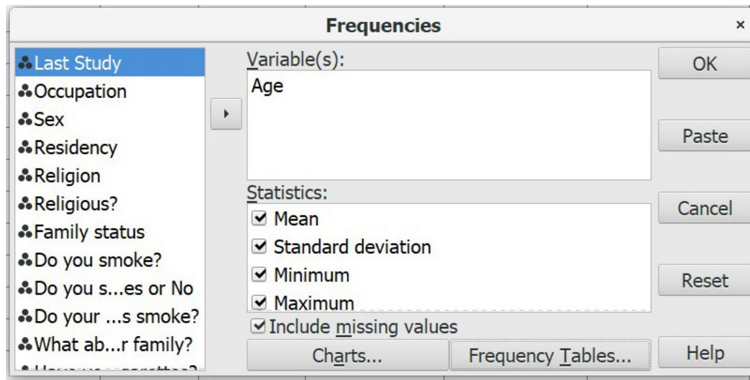
Mivel itt minden változónkat egyenként „górcső” alá vesszük, az adatbázisunk esetleges hibái, elírásai is kiderülnek.

12.4.1. Gyakoriság elemzés és táblázatok

A statisztikai elemzés első lépései közé tartozik a változók elemeinek gyakoriság vizsgálata.

A PSPP és az SPSS parancssora megegyező, a beállítások hasonlóak:
Analyze → Descriptive Statistics → Frequencies

A PSPP-ben a következő beállításokat célszerű eszközölni:



A beállításokat tartalmazó ablakból kiválasztjuk azon változókat, amelyeket elemezni szeretnénk. Jelen példához egy korábbi kutatásom adatbázisát választottam, a *Variable(s)* ablakba egyelőre csak az életkorok (*Age*) címkéjű változót tesszük be. Alatta a *Statistics* ablakban kiválasztjuk azokat a műveleteket, amelyeket el szeretnénk végezni:

Érdemes lekérni az átlagot, standard hibát, minimumot, maximumot, ferdeséget, csúcsosságot, esetleg mediánt (Mean, Standard deviation, Minimum, Maximum, Skewness, Kurtosis, Median), de további lehetőségeink is vannak. Jelöljük be az *Include missing values* opciót is. Amennyiben szeretnénk ábrákat és táblázatokat is kigenerálni, azt a *Charts* és a *Frequency Tables* menüpontok alatt állíthatjuk be.

Most futtassuk le a tesztet és elemezzük a kimeneti ablakot a kísérleti adatbázis alapján:

Age

Value Label	*Value*	*Frequency*	*Percent*	*Valid Percent*	*Cum Percent*
	18	10	3,17	3,21	3,21
	19	23	7,30	7,37	10,58
	20	29	9,21	9,29	19,87
	21	40	12,70	12,82	32,69
	22	43	13,65	13,78	46,47
	23	33	10,48	10,58	57,05
	24	24	7,62	7,69	64,74
	25	31	9,84	9,94	74,68
	26	14	4,44	4,49	79,17
	27	14	4,44	4,49	83,65
	28	13	4,13	4,17	87,82
	29	9	2,86	2,88	90,71
	30	4	1,27	1,28	91,99
	31	2	,63	,64	92,63
	32	2	,63	,64	93,27
	33	1	,32	,32	93,59
	34	3	,95	,96	94,55
	35	2	,63	,64	95,19
	37	1	,32	,32	95,51
	38	1	,32	,32	95,83
	40	1	,32	,32	96,15
	42	1	,32	,32	96,47
	43	2	,63	,64	97,12
	45	1	,32	,32	97,44
	46	1	,32	,32	97,76
	47	1	,32	,32	98,08
	48	1	,32	,32	98,40
	49	1	,32	,32	98,72
	50	1	,32	,32	99,04
	51	1	,32	,32	99,36
	52	1	,32	,32	99,68
	55	1	,32	,32	100,00
	.	3	,95	Missing	
	Total	315	100,0	100,0	

Az táblázat a következő oszlopokat tartalmazza:

Value label: az értékek címkéi, ami jelen esetben azért üres, mert az életkorokhoz nem voltak értékek hozzárendelve a *Variable view* ablakban a változó beállításakor.

Value: a válaszadók konkrét életkora.

Frequency: az egyes életkorok, értékek előfordulásának gyakorisága.

Percent: a válaszok százalékos aránya, beleértve a hiányzó adatokat.

Valid percent: az érvényes válaszok után kiszámított százalékok, tehát a hiányzó válaszok nélkül.

Cum percent: a kumulált százalékok, vagyis az egyes százalékos értékek soronként összeadva.

A *Total* sorban összesíti az egyes oszlopok tartalmát.

A kutatás szempontjából általában az értéket, gyakoriságot és érvényes százalékokat szoktuk megadni. Ne feledjük, ha százalékos arányt írunk szövegtestbe, mindig tegyük oda az eseti számot is a következő formátumban: (példa) a 22 évesek részaránya 13,78% (n=43).

A táblázatból az is látszik, hogy a 30 év felettiek részaránya nagyon alacsony, rendszerint 1-1 esetről van szó, ezért ebben a formában nem fogjuk tudni hasznosítani ezeket az adatokat. A túl alacsony gyakoriság miatt szükséges lesz az életkorokat életkor csoportokba rendezni, így az eredetileg folytonos változónkat ordinális (rangsor) változóvá fogjuk alakítani. Például kialakíthatunk 3 korcsoportot: 18-25 évesek, 25-30 évesek és 30 fölöttiek.

A gyakoriság vizsgálata során a következő táblázatot is megkapjuk a parancssor lefuttatása után:

Age

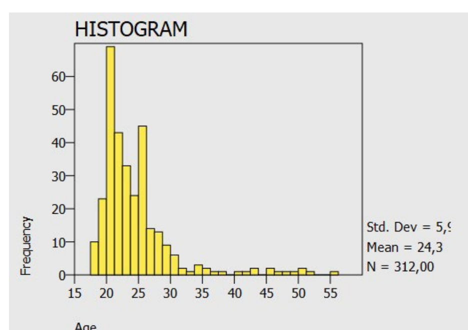
N	*Valid*	312	Érvényes válaszok száma
	Missing	3	A hiányzó válaszok száma
Mean		24,35	Az átlag
Std Dev		5,96	Szórás
Kurtosis		8,27	Ferdeség és csúcsosság
Skewness		2,61	
Minimum		18,00	Minimum és maximum értékek
Maximum		55,00	
Percentiles	50 (Median)	23	Medián

A fenti táblázatból a következő információkat olvashatjuk ki: Az érvényes és hiányzó válaszokat összeadva megkapjuk a minta eseti számát. A mintára jellemző átlag életkor *M*=24,35 év, de a szórás viszonylag magas, *SD*=5,96, ez azt jelenti, hogy az egyes életkorok átlagosan 5,96 évvel térnek el a minta 24,35 éves átlagától. Tehát a mintánk viszonylag heterogén. Az alacsony szórás viszonylag homogén mintára utalna. A ferdeség és csúcsosság, ha nem 0, akkor a változó adateloszlása nem-normális. A minimum és a maximum a mintában előforduló legfiatalabb és legidősebb életkorokat jelöli. A medián (*Md*) az adatsor középső értéke, azaz 23 év.

Az adatok grafikusan is megjeleníthetőek, például diagramokkal is szemléltethetjük őket, valamint táblázatokat is használhatunk (lásd a következő két alfejezetet: 12.4.2. és 12.4.3.).

12.4.2. Az adatok grafikus ábrázolása diagramok segítségével

A PSPP-ben lehetőségünk van 3 típusú grafikon közül választani: hisztogram, oszlopdiagram és kördiagram. Mivel a beállítási lehetőségek és adatvizualizációs opciók a PSPP-ben korlátozottak, javasolt inkább az Excel vagy SPSS használata, esetleg egyéb, grafikonok készítésére alkalmas szoftverek. Az alábbi példákból evidenssé válik, hogy miért nem javasolt a PSPP használata grafikonok rajzolásához.



A fenti ábrán egy PSPP által kigenerált hisztogramot látunk. A hisztogram, mint adatmegjelenítési forma akkor használható, ha egy csoportra vonatkozó meghatározott tartományon belüli statisztikai adatot szeretnénk ábrázolni. Folytonos típusú változók megjelenítésére alkalmas, ezért a sávelemek (oszlopok) átrendezése nem lehetséges, mindig számsorrendben kell megjelenjenek a legalacsonyabbtól a legmagasabbig. Nagyszámú adatot és ezek eloszlását jeleníti meg, melyek ötféle eloszlás jellemezhet:

a. normális eloszlás: a hisztogram pontjai egy harang alakú, szimmetrikus görbét jelenítenek meg.

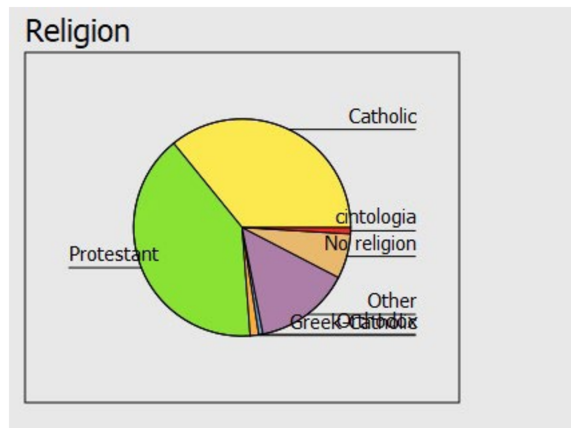
b. bimodális eloszlás: két különálló csúcsú hisztogramok leírására használják; olyankor fordul elő, amikor az adathalmaz két különböző csoportosítással kapcsolatos megfigyeléseket tartalmaz.

c. Jobbra ferde eloszlás: a hisztogram adatainak nagy része a csúcsának bal oldalán található.

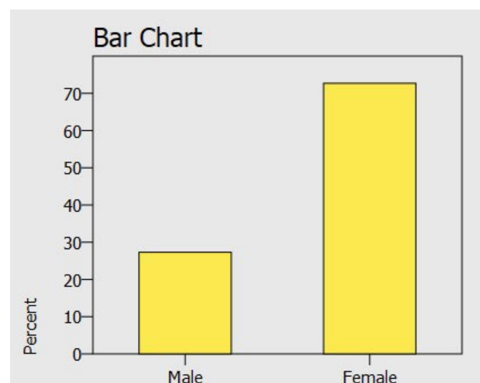
d. Balra ferde eloszlás: az adatok nagy része a hisztogram csúcsának jobb oldalán található.

e. Véletlenszerű eloszlás: ha nincs észrevehető mintázat és több csúcsot is tartalmaz a hisztogram.

A hisztogramok megjelenítése is több formában lehetséges, például oszlopokkal, pontokkal, összekötött pontokkal, vonallal, terület alapú elhatárolással.



A fenti ábrán egy PSPP által kigenerált kördiagramot láthatunk. Ez sajnos nem is szép és nem is eléggé informatív, így ebben a formában nem használható a dolgozatban; a százalékos értékeket is meg kellene jeleníteni a szeletekben. Ez a „pite” vagy „sajt” típusú ábrázolás kategorikus adatokhoz ajánlott, ahol a változó minden kategóriája egy-egy szeletnek felel meg. Amint a fenti példából is jól megfigyelhető, ez a típusú ábrázolás nem alkalmas 5-6 kategóriánál több megjelenítésére, mivel túlzsúfolttá és ezáltal nehezen értelmezhetővé válik. A kördiagram az adott kategóriának az egészhez viszonyított százalékos arányát tükrözi (tehát a szeletek mérete relativizált). A szeletek helye tetszőlegesen felcserélhető. Hátránya, hogy a kördiagramon csak egyetlen változó jeleníthető meg.



A fenti oszlopdiagramot szintén a PSPP-vel generáltuk ki. Itt is szembetűnik egy hiányosság, mivel a konkrét százalékos vagy gyakorisági értékek nem látszanak külön-külön az oszlopok tetején vagy az oszloboknan. Oszlopdiagramot olyankor ajánlott használni, amikor két tengelyen ábrázolható kategorikus adatot jelenítünk meg. A diagramok által használt mértékegység a leggyakrabban a darabszám (gyakoriság) vagy a százalékos arány, esetleg összegek. Az oszlopdiagram a kategóriákat egymáshoz viszonyítva jeleníti meg, a váltzó kategóriáinak abszolút értékeit ábrázolja. A diagram megjelenítése lehet horizontális vagy vertikális, illetve csoportosított, amikor több változót ábrázolunk ugyanazon a

diagramon. A trendek és mintázatok megjelenítésére is alkalmas. Előnyük, hogy sok kategóriát tartalmazó adatszettek is ábrázolhatóak az oszlopdiagramok segítségével. Hátrányuk, hogy a komplexebb diagramok nehezen értelmezhetőek, ezért szöveges magyarázatra szorulnak.

Általánosan érvényes, hogy nem minden adatot szükséges grafikus formában megjeleníteni, csak azt, amit hangsúlyozni szeretnénk. Ugyanakkor a grafikusan megjelenített adatokat mindig meg kell említeni a szövegtestben, utalva a diagram vagy ábra számára. A magyarázatok inkább a komplexebb, például több változót is megjelenítő oszlopdiagramjainknál legyenek részletesebbek, ami könnyen értelmezhető ábra, ott csak a számunkra leglényegesebb adatokból levonható következtetést írjuk le. Gyakori hiba, hogy a hallgató teletűzdeli egyszerű „sajtokkal" a dolgozatát, nem ír hozzá magyarázatokat, vagy ha ír is, csak elnagyoltan. Nincs arra konkrét recept, hogy hány diagram lehet egy dolgozatban, de igyekezzünk felváltva használni a grafikonokat a táblázatokkal. Figyeljünk arra, hogy a dolgozat ne keltsen olyan benyomást, mintha az érdemi munka helyett csak grafikus elemekkel lenne kitöltve.

12.4.3. Táblázatok és kereszttáblák

Az adatok vizualizációjának gyakran használt módja a táblázatok szerkesztése. A táblázatokban megjeleníthető, hogy egy bizonyos érték hányszor fordul elő a mintában. Kifejezhetjük eseti számban és/vagy százalékos arányban. A százalékos arány feltüntetése önmagában nem elegendő, az esetek számát is mindig be kell írni, mint pl. 72% (*n*=28). Ebben a példában a 28 az érvényes válaszadók számát jelöli.

A táblázatoknak tartalmazniuk kell az érvényes válaszok számát, a hiányzó válaszok számát, illetve az összesítésüket is.

Kétféle táblázatot szerkeszthetünk: egyszerű táblázatokat, ahol csak egy változót jelenítünk meg, annak minden kategóriájával, illetve kereszttáblákat, ahol több változót van lehetőségünk szimultán összevetni.

Az alábbiakban egy korábbi kutatásomból hoznék fel egy példát egyszerű gyakorisági táblázatra, illetve a hozzá társított magyarázatra:

A 196 érvényes válasz alapján az egyetemi képzést a segítőszakemberek közepes és közepesnél kicsivel jobbnak tartják (63,3%, sz=148) – 1. sz. táblázat. A képzéssel elégedetlenek összesített részaránya 13,7% (sz=32), teljesen elégedett a képzéssel mindössze a válaszadók 6,8%-a (sz=16).

1. sz. Táblázat – Mennyire elégedett az egyetemi képzéssel?

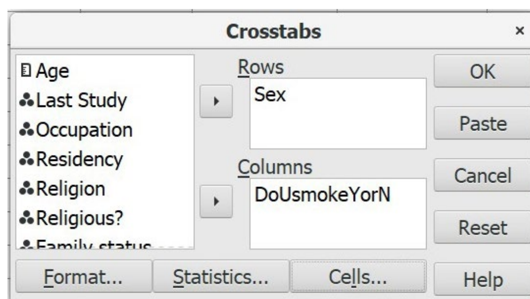
		Gyakoriság	Százalék	Érvényes százalék
Érvényes válaszok	Egyáltalán nem	21	9.0	10.7
	Kismértékben	11	4.7	5.6
	Közepesen	83	35.5	42.3
	Elég jó	65	27.8	33.2
	Teljesen jó	16	6.8	8.2
	Összesen	196	83.8	100.0
Hiányzó válaszok		38	16.2	
Összesen		234	100.0	

(Forrás: Albert-Lőrincz, Belényi, 2019, 6.)

Figyeljünk arra, hogy a táblázatokat minden esetben vagy megelőznie, vagy követnie kell egy kísérő-magyarázó szövegrésznek, mely összefoglalja, hogy az adott táblázatban mit látunk. Fejezetet soha ne zárjunk táblázattal, egy-két témából kivezető vagy témára átvezető mondatra minden esetben szükség van. A táblázatok APA hivatkozási stílus szerinti formai követelményeit a 6. fejezetben ismertettük.

A kereszttáblák kettő vagy több változó közötti kapcsolatot szemléltetik. A kereszttáblákban általában az egyik változónk független (pl. nemek, csoportok), a többi függő változó (pl. lehetnek attitűdök, szokások, preferenciák, ismeretek stb.).

A **PSPP** és az **SPSS** parancssora megegyező, a beállítások hasonlóak:
Analyze → Descriptive Statistics → Crosstabs



A beállításoknál a változóink közül kiválasztjuk, hogy mit szeretnénk a táblázat soraiban és mit szeretnénk a táblázat oszlopaiban megjeleníteni. Fontos beállítások a *Statistics* fül alatt találhatóak, itt azokat a teszteket állíthatjuk be, amelyeket szeretnénk lefuttatni a táblázatban szereplő változóinkon (például Khi-négyzet teszt, korreláció stb.). A *Cells* fülnél azt állíthatjuk be, hogy az összegzést soronként és/vagy oszloponként szeretnénk látni, mik az elvárt érékek, maradvány értékek és determinációs együttható[17]. A kereszttáblák változóival összefüggésbe hozható próbákra a következő alfejezetben visszatérek.

17 Egyezményes jele az R^2, a függő változó varianciájának a független változóból megjósolható arányát fejezi ki.

Lehetőség szerint igyekezzünk minél átláthatóbb táblázatokat készíteni, ennek megfelelően korlátozzuk a változóink számát. Például a független változóinkat több táblázatba is lebonthatjuk, amennyiben a független változóinkat valamilyen szempont alapján csoportosítani tudjuk.

Az alábbiakban egy PSPP-ben készített kereszttáblát láthatunk. A *Cells* fülnél a *Cell display* ablakban a *Count* és *Row* opciókat jelöltük be.

Sex * Do you smoke? Yes or No [count, row %].

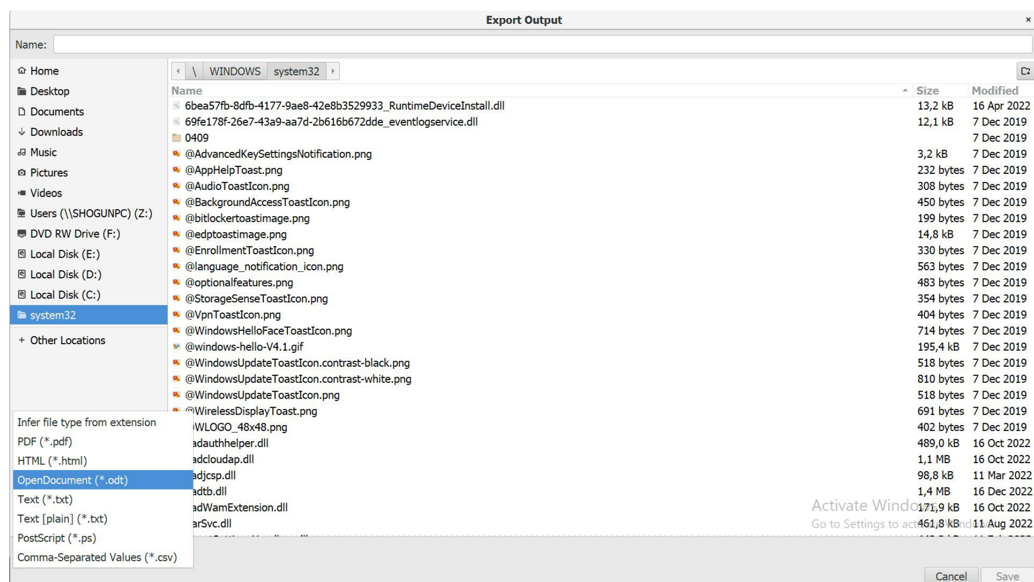
	Do you smoke? Yes or No		
Sex	No	Yes	Total
Male	32,00 37,21%	54,00 62,79%	86,00 100,00%
Female	85,00 37,12%	144,00 62,88%	229,00 100,00%
Total	117,00 37,14%	198,00 62,86%	315,00 100,00%

A fenti példában soronként megjelenítésre kerül a gyakoriság és százalékos előfordulási arány. Az adatokat elemezve kitűnik, hogy ebben a mintában a nők felülreprezentáltak (315 válaszadóból 229 nő és csupán 86 férfi), valamint, hogy a két csoporton belül szinte azonos százalékos arányban fordulnak elő a dohányzó és nem dohányzó alanyok.

A csoportok közötti hasonló gyakoriság és százalékos arány azonban még nem a teljes kinyerhető eredmény, ebből még nem tudjuk, hogy van-e kapcsolat a változók között. A megfelelő tesztet a változóink típusának függvényében választjuk ki (jelen példában két kategorikus változónk van, ezért Khi-négyzet (egyezményes jele: X^2) függetlenségpróbát végezhetünk)[18].

A táblázatokat a PSPP kimeneti ablakából egy többlépcsős módszerrel emelhetjük át a dolgozatunkba. Sajnos a PSPP felülete az SPSS-el ellentétben nem teszi lehetővé a táblázatok, grafikonok copy-paste alapú másolását. A kimeneti ablak *File* menüpontjából kiválasztjuk az *Export* parancsot. Figyeljünk arra, hogy nem az adatbázis nézet menüpontját, hanem a kimeneti ablak (*Output window*) menüpontját keressük. Ha helyesen jártunk el, az alábbi *Export output* ablak nyílik meg:

18 Lásd a 3. sz. Táblázatot a 12.5.2. alfejezetben.



Az ablak tetején a *Name* sorba beírjuk a tetszőleges fájlnevet, kiválasztjuk a célkönyvtárat (mappát), majd a bal alsó sarokból a kívánt fájl-formátumot. Javasolt az **.odt* Open Office által is ismert formátumot választani, mivel ezt a *Microsoft Office* is ismeri. Végül a jobb alsó sarokban található *Save* fülre kattintva elmentjük.

Ezután a célkönyvtárból az általunk használt szövegszerkesztővel megnyitjuk a kimeneti eredményeinket tartalmazó fájlt és szabadon szerkeszthetjük. Esetleg átvihetjük Excelbe is, ezáltal további szerkesztési és grafikus ábrázolási lehetőségeket érhetünk el.[19]

12.5. Nem paraméters tesztek

12.5.1. Pearson-féle Khi négyzet (χ^2) függetlenség próba

A *Khi négyzet próba* csoportok közti kapcsolat vizsgálatára alkalmas. A kereszttáblákból megtudhatjuk az egyes változók gyakoriságát és százalékos arányát, de ebből még nem tudjuk, hogy van-e köztük kapcsolat. A változóinkat megpróbáljuk értelem és a kutatás célja szerint függő és független felosztásban értelmezni, a beállításokkor a független változó a sorba, a függő változó az oszlopba kerül.

19 A táblázatokat érdemesebb Excelben formázni és véglegesíteni, mert rugalmasabb és több lehetőséget is biztosít, mint a Word szövegszerkesztő. A táblázatokban szereplő angol nyelvű kifejezéseket is szükséges lefordítani a szakdolgozat nyelvére.

Feltételei:

– A művelethez két diszkrét, kategorikus változóra[20] van szükségünk, kettő vagy több kategóriával. Az elvárt gyakoriság a kereszttábla összes cellájának kevesebb mint 20%-ában lehet kisebb, mint 5.[21] Amennyiben ez a feltétel nem teljesül, vagy nagyon kicsi a minta, a 2x2-es kontingencia táblázatoknál a *Fisher féle egzakt teszt* (Fisher's Exact Test) szerinti eredményt kell figyelembe venni[22].

Parancssor és beállítások:

Az **SPSS** és **PSPP** parancssora megegyezik, a beállítások hasonlóak.

Analyze → Descriptive Statistics → Crosstabs → Statistics → Chi-square

A százalékokat a független változó szerint, vagyis soronként jelenítsük meg. A *Cells* fülnél a *Count* és *Row* opciókat jelöljük be.

A vizsgálati hipotézis:

H_0 – a változók között nincs összefüggés.

H_1 – a változók között van összefüggés.

Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns ($p>.05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a változók között nincs összefüggés.

Ha a teszt eredménye szignifikáns ($p<.05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a változók között van összefüggés.

Magyarázat egy példán keresztül:

Summary.

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Religious? * Do you smoke? Yes or No	315	100,0%	0	0,0%	315	100,0%

Religious? * Do you smoke? Yes or No [count, row %].

	Do you smoke? Yes or No		
Religious?	No	Yes	Total
Yes	65,00 43,05%	86,00 56,95%	151,00 100,00%
No	36,00 28,35%	91,00 71,65%	127,00 100,00%
I don't know	16,00 43,24%	21,00 56,76%	37,00 100,00%
Total	117,00 37,14%	198,00 62,86%	315,00 100,00%

Chi-square tests.

Statistic	*Value*	*df*	*Asymp. Sig. (2-tailed)*
Pearson Chi-Square	7,05	2	,029
Likelihood Ratio	7,17	2	,028
Linear-by-Linear Association	1,29	1	,256
N of Valid Cases	315		

20 A teszt akkor is elvégezhető, ha folytonos változókból alakítunk ki kategorikus változókat, vagyis egy adott szempont alapján csoportokat alakítunk ki.

21 Az SPSS a kimeneti ablakban a táblázat alján jelzi, ha ez a feltétel nem teljesült, a PSPP sajnos nem, ezért nekünk kell ellenőriznünk. Ha nem teljesül a feltétel, lehet próbálkozni a változó kategóriáinak átcsoportosításával (pl. összevonásával), vagy másik tesztet kell végezni.

22 2x2-es kontingencia táblázatok esetén nem szükséges egyéb beállítás, a Fisher egzakt teszt eredményét is automatikusan kiszámolja a PSPP. Bár elvileg lehetséges nagyobb táblázatokban is alkalmazni, általában a 2x2-es táblázatokhoz szokták használni.

A példánkban két kategorikus változót vizsgáltunk. A kereszttáblában a vallásosságot tekintettük független változónak, függőnek azt, hogy az alanyok dohányoznak-e. Jelen esetben mi dönthetjük el, hogy melyik változót tekintjük függetlennek. A kérdésünk az, hogy a vallásosság és a dohányzás káros szokása között van-e kapcsolat. Mivel a Khi-négyzet teszt eredménye szignifikáns, (p=.029 azaz p<.05), megállapíthatjuk, hogy a vallásosság és a dohányzás között van kapcsolat. A teszt nem alkalmas ok-okozati összefüggések kimutatására, azt azonban megállapíthatjuk, hogy a mintánkban szereplő alanyok többsége bár dohányzik, a dohányzási szokás aránya kisebb azok között, akik vallásosnak tartják magukat. A khi-négyzet a kapcsolat erősségét, vagyis a hatás nagyságát nem adja meg, de ha szeretnénk kiszámolni, akkor alkalmazhatjuk a *Phi* és *Cramer-V* teszteket[23]. Ha a mintánk reprezentatív, akkor ezt a megállapítást kiterjeszthetjük a teljes populációnkra.

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

Mivel a Khi-négyzet teszt eredménye szignifikáns, megállapíthatjuk, hogy a vallásosság és a dohányzás között van kapcsolat (X^2 (2, n = 315) = 7.05, p < .05).

Jelölése:

$$\boldsymbol{X^2\ (2,\ n = 315) = 24.12,\ p < .05}$$

ahol Khi-négyzet (szabadsági fokok (*df*), esetek száma (*n*) = a teszt eredménye, *p* érték

12.5.2. Khi-négyzet illeszkedésvizsgálat

A *Khi-négyzet illeszkedésvizsgálat* egy kategorikus változó értékeinek eloszlását (arányulását) fejezi ki a változó specifikus részeihez képest, vagy egy elvárt értékekhez képest. A teszt abból a hipotézisből indul ki, hogy a változónk értékeinek eloszlása arányaiban nem különbözik egy elvárt 50-50%-os arányulástól, vagy egy tetszőlegesen beállított arányulástól. Ez utóbbi lehetőséget ad a változónk összevetéséhez egy korábbi vagy másik vizsgálatból, tanulmányból nyert arányhoz képest.

Feltételei:

– a művelethez egy diszkrét, kategorikus változóra[24] van szükségünk, kettő vagy több kategóriával;

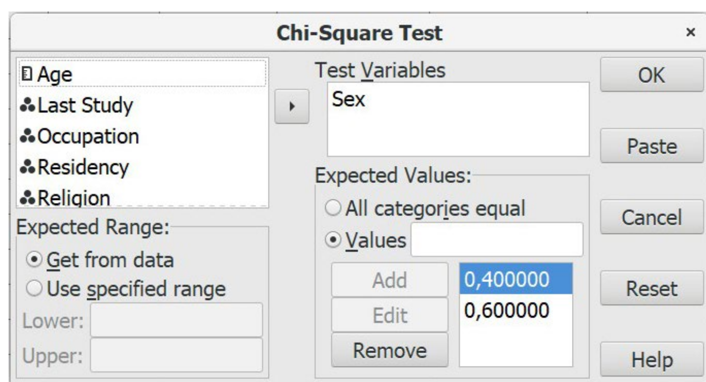
23 A 2x2-es táblázatoknál a Phi együttható a mérvadó, az ennél nagyobb táblázatoknál Cramer-V féle korrelációt számítunk. Az együtthatók értéke 0 és 1 között mozoghat, ahol Pallant (2007) szerint a 0 közeli értékek gyenge, a 0.3 körüli értékek közepes, a 0.5 fölötti értékek erős összefüggésre utalnak, viszont Haldun (2018) szerint már a 0.25-ös érték is erős összefüggésre utal.

24 A teszt akkor is elvégezhető, ha folytonos változókból alakítunk ki kategorikus változókat, vagyis egy adott szempont alapján csoportokat alakítunk ki.

– szükségünk van egy elvárt arányulásra is (pl. férfiak 40%, nők 60%, amit numerikusan .4/.6 formátumban fejezhetünk ki).

Parancssor és beállítások:

Az **SPSS** és **PSPP** parancssora különbözik, de a beállítások szinte azonosak.
SPSS: Analyze → Non-Parametric Tests → Legacy Dialogs → Chi-square
PSPP: Analyze → Non-Parametric Statistics → Chi-square



A *Test Variables* ablakba betesszük a vizsgálni kívánt változót, jelen esetben a nemet. A beállításoknál lehetőségünk van megadni egy elvárt tartományt, amennyiben ezt nem ismerjük, akkor hagyjuk az alapbeállításon, hogy a teszt ezt a változónkból számítsa ki (*Expected Range - Get from data*). Az elvárt értékeknél az arányulást adjuk meg, az alapbeállítás (*Expected Values – All categories equal*) szerint a változónk kategóriáinak 50-50%-os eloszlása lenne (ha a változónk több kategóriát tartalmaz, akkor ez a kategóriák számának megfelelően automatikusan változik). A másik lehetőség, amit a példában is alkalmaztunk, hogy megadjuk az elvárt arányulást, mely szerint a mintánkban 40% férfi és 60% nő volna az elvárt érték. Ezt az értékeknél állítjuk be, 0.4 (*Expected Values – Add 0.4 és Add 0.6*). Ahány kategóriánk van, annyi arányulási értéket kell megadnunk. Ezen értékeket célszerű korábbi vizsgálatok vagy egyéb tanulmányok alapján megadni. A példa kedvéért tételezzük fel, hogy ismerjük egy teljes populáció nemek szerinti eloszlását, ami történetesen 40-60%. Ha azt szeretnénk megvizsgálni, hogy ehhez képest a mintánk nemekre vonatkozó változója hogyan arányul, akkor a khi-négyzet illeszkedésvizsgálat lehet a megfelelő teszt.

A vizsgálati hipotézis:

H_0 – a változó értékeinek eloszlása megegyezik az elvárt értékek eloszlásával.

H_1 – a változó értékeinek eloszlása különbözik az elvárt értékek eloszlásától.

Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns ($p>.05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a változónk arányulása megegyezik az elvárt arányulással.

Ha a teszt eredménye szignifikáns ($p<.05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a változók arányulása különbözik az elvárt arányuláshoz képest.

Magyarázat egy példán keresztül:

Ha a fenti beállításokat lefuttatjuk a példánk kedvéért az adatbázisunkon, a következő eredményt kapjuk a PSPP kimeneti ablakában:

```
NPAR TEST
     /CHISQUARE= Sex
     /EXPECTED =  0.4 0.6.
```

Sex

	Observed N	Expected N	Residual
Male	86	126,00	-40,00
Female	229	189,00	40,00
Total	315		

Test Statistics

	Sex
Chi-Square	21,16
df	1
Asymp. Sig.	,000

Láthatjuk, hogy a mintában 86 férfi és 229 nő szerepelt, összesen n=315 alany. Ez az u.n megfigyelt (*Observed N*) érték. Az általunk megadott elvárt 40-60 százalékot a program az össz-alanyszámból számítja ki. Jelen esetben 315*40/100=126, tehát ahhoz, hogy az elvárt 40%-os arányuláshoz teljesüljön, 126 férfinak kellene szerepelnie a mintában (*Expected N*), de nekünk csak 86 alanyunk van, azaz enynyi a megfigyelt érték. Ez elég nagy különbség, amit a teszt eredménye is alátámaszt, ugyanis szignifikáns (*Test Statistics* táblázat). Így elvetjük a nullhipotézist és kijelenthetjük, hogy a mintánk arányulása különbözik az elvárt értéktől.

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

A mintánk arányulása különbözik az elvárt értéktől
(X^2 (1, n = 315) = 21.16, p < .01).

Jelölése:

$\mathbf{X^2\ (1,\ n = 315) = 21.16,\ p < .01}$

ahol Khi-négyzet (szabadsági fokok (*df*), esetek száma (*n*)) = a teszt eredménye, *p* érték

12.5.3. McNemar féle teszt

A *McNemar teszt*tel azonos alanyok két időpontban felvett válaszai közötti kapcsolatot vizsgálhatunk. Alkalmazható két kategorikus kétértékű (bináris) változó eloszlása közötti kapcsolat vizsgálatára. Ismérve a 2x2-es kontingencia mátrix. Gyakori alkalmazása a két különböző időpontban felvett mérések (pre- és poszttesztek) eredményeinek összehasonlítása. A teszt arra ad választ, hogy az alanyok válaszai arányaiban különböznek-e a két mérési időpontban.

Feltételei:

– Két kategorikus változó, két megegyező válaszlehetőséggel (pl. igen/nem; igaz/hamis).

– Mindkét változó ugyanarra az információra vonatkozik, ugyanazon alanyoktól (mintától) származik, csak különböző időpontokban.

Parancssor és beállítások:

Az **SPSS** parancssora megegyezik a kereszttáblákéra vonatkozóval; a McNemar tesztet is be lehet jelölni. **Analyze → Descriptive Statistics → Crosstabs → Statistics**

A **PSPP** grafikus felhasználói felülete[25] nem tartalmazza a teszthez szükséges parancssort, azonban a teszt beprogramozható[26]. Nyissuk meg a *Syntax Editor* ablakot:

File→New→Syntax

Kétsoros parancssort kell beírnunk:

NPAR TEST

/MCNEMAR *változó1* WITH *változó2* (PAIRED)

A "*változó1*" és "*változó2*" helyett az adatbázisunkban használt kategorikus változók nevét adjuk meg.

Majd a *Run* menüpontból lefuttatjuk, az *All* menüpontra kattintva. Az eredményeket a kimeneti ablakban találjuk.

A vizsgálati hipotézis:

H_0 – a minta arányaiban nem különbözik a két mérési időpontban.

H_1 – a minta arányaiban különbözik a két mérési időpontban.

25 A PSPP grafikus felhasználói felülete kevesebb lehetőséget tüntet fel a legördíthető menük és műveletek között, mint amennyi parancssor kódolással elvégezhető.

26 A teszt beállításairól bővebben lásd a PSPP kezelési útmutató 15.11.9-es szekcióját. A PSPP kezelési útmutatója a következő linken érhető el: https://www.gnu.org/software/pspp/manual/pspp.pdf

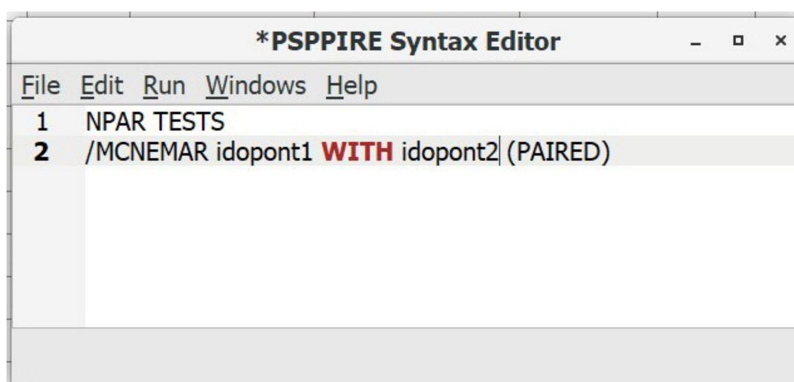
Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns ($p > .05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a minta arányaiban nem különbözik a két mérési időpontban.

Ha a teszt eredménye szignifikáns ($p < .05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a minta arányaiban különbözik a két mérési időpontban.

Magyarázat egy példán keresztül

Tételezzük fel, hogy tanulók egy csoportja (n=20) két tesztet ír egymás utáni időpontokban. A teszteknek két kimenetele lehet: vagy teljesítik vagy nem (ez egy bináris vagy más néven dichotóm változó). A PSPP *Syntax Editorában* beállítjuk a McNemar tesztet az alábbiak szerint:



A teszt lefuttatása után (*Run→All*) a következő eredményeket kapjuk a kimeneti ablakban:

NPAR TESTS

Első teszt & Második teszt

Első teszt	Második teszt	
	teljesítette	nem teljesítette
teljesítette	4	5
nem teljesítette	9	2

Test Statistics

	N	*Exact Sig. (2-tailed)*	*Exact Sig. (1-tailed)*	*Point Probability*
Első teszt & Második teszt	20	1,820	,910	,12

A *McNemar próba* eredménye nem szignifikáns ($p = 1.82$), tehát fenntartjuk a nullhipotézist. Mivel a próba eredménye nem szignifikáns ($p > .05$), arra következtethetünk, hogy a tanulók ($n = 20$) által a két időpontban megírt tesztek eredményei arányaikban nem különböznek egymástól.

Az eredmények megjelenítése a dolgozatban:

Leírjuk, hogy *McNemar teszt*tel vizsgáltuk a mintán belüli csoportok közötti kapcsolatot két mérési időpontban; az eredmény szignifikáns-e vagy sem; illetve ennek megfelelően különbözik-e vagy sem arányaiban a minta a két mérési időpontban. A szignifikancia szint mellett megadjuk az esetek számát (*n*) és a *p* értéket is.

Jelölése:

n = 20, $p > .05$

ahol *n* az eseti szám, *p* a szignifikancia szint.

12.5.4. Cohran féle Q teszt

A *Cohran féle Q teszt* hasonlít a McNemar teszthez, viszont ebben az esetben az alanyok válaszainak három vagy több különböző időpontban felvett méréseinek eloszlása közötti kapcsolatot vizsgálhatjuk. Alkalmazható három vagy több kategorikus kétértékű (bináris) változó eloszlása közötti kapcsolat vizsgálatára. A teszt arra ad választ, hogy az alanyok válaszai arányaiban különböznek-e a három vagy háromnál több mérési időpontban. A *Cohran féle Q teszt* kiszámolja, hogy van-e különbség, de azt nem, hogy hol, mely időpontok között vannak ezek a különbségek. Ha a páronkénti különbségeket szeretnénk kiszámolni, akkor McNemar tesztet alkalmazhatunk.

Feltételei:

– Három vagy több kategorikus változó, két megegyező válaszlehetőséggel (pl. igen/nem; igaz/hamis).

– Mindegyik változó ugyanarra az információra vonatkozik, ugyanazon alanyoktól (mintától) származik, csak különböző időpontokban.

Parancssor és beállítások:

Az SPSS és PSPP parancssora különbözik, de a beállítások szinte azonosak.

SPSS: Analyze → Nonparametric Tests → Legacy Dialogs → K Related Samples

A változók beállítása után a *Test Type* ablakban kivesszük az alapértelmezett *Friedman* tesztet és bejelöljük a *Cochran's Q* opciót.

PSPP: Analyze → Non-Parametric Statistics → K Related Samples

A *Test Type* opciói közül kiválasztjuk a *Cochran's Q* opciót és kivesszük az alapértelmezett *Friedman* tesztet.

A vizsgálati hipotézis:

H_0 – a minta arányaiban megegyezik a három (vagy több) mérési időpontban.
H_1 – a minta arányaiban különbözik a három (vagy több) mérési időpontban.

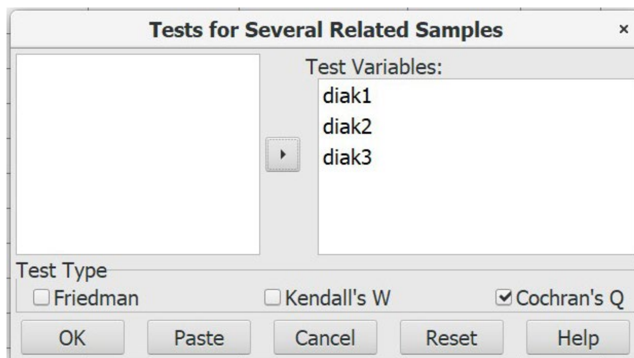
Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns (p>.05), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a változók arányaikban megegyeznek a mérési időpontokban.

Ha a teszt eredménye szignifikáns (p<.05), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a változók arányaikban különböznek a mérési időpontokban.

Magyarázat egy példán keresztül:

A példánk alapjául szolgáló adatbázisban 20 diák szerepel, akik a félév során három tesztet írtak. Az eredményeket nem pontoztuk, hanem a teljesített, vagy nem teljesített kategóriákba soroltuk (bináris változók). A parancssor megadása után a *Test Variables* ablakba betesszük a három változót, melyek a diákok különböző időszakokban írt tesztjeinek eredményeit tartalmazzák. A beállításoknál kivesszük az alapértelmezett *Friedman teszt*et és helyette bejelöljük a *Cohrans's Q* opciót.



A fenti beállítások szerint lefuttatott teszt az alábbi eredményt jeleníti meg a kimeneti ablakban:

NPAR TESTS

```
NPAR TEST
    /COCHRAN =  diak1 diak2 diak3.
```

Frequencies

	Value	
	Success (2)	Failure (1)
Első teszt	9	11
Második teszt	9	11
Harmadik teszt	12	8

Test Statistics

N	20
Cochran's Q	1,06
df	2
Asymp. Sig.	,589

A gyakoriság táblázatban (*Frequecies*) azt figyelhetjük meg, hogy az első két teszt esetében azonos eredményeket kaptunk, a harmadik teszt azonban több diáknak sikerült. Erről a különbségről még nem tudjuk, hogy statisztikailag szignifikáns-e. A *Test Statistics* táblázatban láthatjuk, hogy a $p>.05$ ($p=.589$) érték nem szignifikáns, tehát fenntartjuk a nullhipotézist és megállapítjuk, hogy a változók arányaikban megegyeznek a mérési időpontokban. Ez az eredmény a három mérési időpontra vonatkozik. Ebből arra következtethetünk, hogy a diákok a félév során a teszt eredményeinek arányát tekintve nem fejlődtek (annak ellenére, hogy az utolsó teszten többen átmentek, mint az előző kettőn). Amennyiben azt szeretnénk megvizsgálni, hogy a második és harmadik teszt között van-e szignifikáns különbség, alkalmazhatjuk a *McNemar féle tesztet.*

Az eredmények megjelenítése a dolgozatban:

A diákok különböző időszakokban írt tesztjeinek eredményeit *Cohran féle Q próbá*val hasonlítottuk össze. Mivel a próba nem jelzett statisztikailag szignifikáns eltérést, arra következtethetünk, hogy a három mérési időpont eredményei arányaikban nem különböznek ($CQ\,(2, n = 20) = 1.06, p > .01$).

Jelölése:

$$\boldsymbol{CQ\,(2, n = 20) = 1.06, p>.01}$$

ahol Cohran Q teszt (szabadsági fokok (*df*), esetek száma (*n*)) = a teszt eredménye, *p* érték

12.5.5. Cohen féle kappa együttható

A *Cohen-féle kappa együttható* (röviden κ együttható) egy olyan asszociációs mérőszám, amely két kategorikus változó egyezését vizsgálja, amikor kódolásuk eltérő.[27] A nominális értékeléseket tartalmazó változók két külön megfigyelőnek, bírálónak vagy értékelőnek nézetkülönbségét vagy nézetazonosságát adja meg. A teszt kiszámolja, hogy a kétféle kódolás különbözik-e egymástól (nézetkülönbség), vagy az osztályozás egybehangzó (nézetazonosság). A κ együttható sokféle alkalmazása elképzelhető, gyakran használják például tesztek validitásának vizsgálatára (amennyiben ugyanazt a jelenséget mérik).

Feltételei:

– Két kategorikus változó, amelyeknek azonos számú kategóriája van.
– A változók kódolása különbözik.

27 Az eltérő kódolás eredhet a bíráló szubjektív megítéléséből is.

Parancssor és beállítások:

Az SPSS és PSPP parancssora megegyezik, a beállítások szinte azonosak.

Analyze → Descriptive Statistics → Crosstabs → Statistics → Kappa

A vizsgálati hipotézis:

H_0 – az értéklések között nézetkülönbség van, vagy a kétféle kódolás egymástól független.

H_1 – az értékelések között nézetazonosság van, vagy a kétféle kódolás egybehangzó.

Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns (p>.05), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy az értékelések egymástól függetlenek.

Ha a teszt eredménye szignifikáns (p<.05), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a értékelések egybehangzóak.

A minta nagysága befolyásolja, hogy a statisztikailag szignifikáns kappa együttható kimutatható legyen, ami több tényezőtől is függ, mint például a kategóriák számától, az egyező válaszok arányától vagy a teszt kitűzött statisztikai erejétől[28] (Sim és Wright, 2005). A szükséges mintaméret pontosan számolható (Mohamad és Nur, 2017), 25-50 között már megbízható eredményt kaphatunk, de az említett szerzők által közölt táblázatok szerint esetenként akár kisebb minta is elegendő.

A kappa-mutató értékei általában 0-tól 1-ig terjedhetnek, ahol “0” az egyezés hiánya, “1” a tökéletes egyezés[29]:

0-tól 0,2-ig: nagyon gyenge;

0,2-től 0,4-ig: gyenge;

0,4-től 0,6-ig: közepes;

0,6-tól 0,8-ig: jó;

0,8-tól 1-ig kiváló.

A κ együttható értékei a tizedes nagyságrendben eltérő besorolással is előfordulnak a szakirodalomban, de alapjában véve a fenti besorolás elfogadható.

Magyarázat egy példán keresztül:

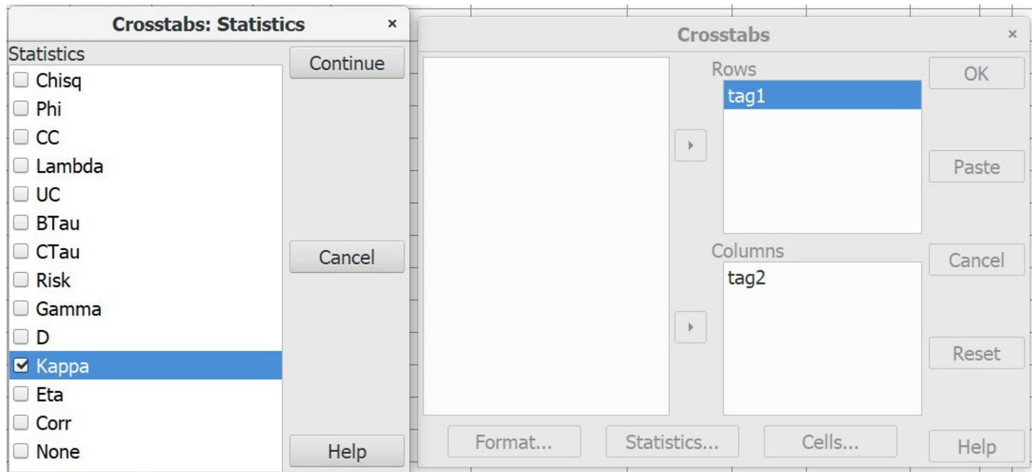
Képzeljünk el egy többfordulós állásinterjút, ahol két bizottsági tag 20 jelentkező közül döntheti el, hogy kit hívnak be, kit szeretnének ismét meghallgatni és kit nem alkalmaznának biztosan. Tehát az értékelés három kategóriát tartalmaz: továbbjut / újra behívják / nem alkalmazzák). A κ együtthatóval ki-

28 A statisztikai erősség annak valószínűségére utal, hogy a szignifikancia-teszt a nullhipotézistől való eltérést észleli, vagyis a nullhipotézist helyesen veti el. Általában elfogadott a 80 vagy 90%-os szint.

29 Elvileg negatív értéket is felvehet, ami olyan nézetkülönbségre utal, amely a véletlenszerűen elvártnál is nagyobb (Sim és Wright, 2005).

számítható, hogy az interjú alanyok megítélésében két bizottsági tag között van-e nézetazonosság.

A két bizottsági tag válaszait tartalmazó változókat tetszőleges sorrendben betesszük a sorok és oszlopok ablakba.



A teszt lefuttatása után a kimeneti ablakban az alábbi eredményeket kaptuk:

```
CROSSTABS
     /TABLES= tag1 BY   tag2
     /FORMAT=AVALUE TABLES PIVOT
     /STATISTICS=KAPPA
     /CELLS=COUNT ROW.
```

Summary.

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Bizottsági tag 1 * Bizottsági tag 2	20	100,0%	0	0,0%	20	100,0%

Bizottsági tag 1 * Bizottsági tag 2 [count, row %].

	Bizottsági tag 2			
Bizottsági tag 1	továbbjut	újra behívják	nem alkalmazzák	Total
továbbjut	5,00 83,33%	1,00 16,67%	,00 ,00%	6,00 100,00%
újra behívják	1,00 12,50%	5,00 62,50%	2,00 25,00%	8,00 100,00%
nem alkalmazzák	1,00 16,67%	,00 ,00%	5,00 83,33%	6,00 100,00%
Total	7,00 35,00%	6,00 30,00%	7,00 35,00%	20,00 100,00%

Symmetric measures.

Category	*Statistic*	*Value*	*Asymp. Std. Error*	*Approx. T*	*Approx. Sig.*
Measure of Agreement	Kappa	,63	,14	4,02	
N of Valid Cases		20			

A táblázat megadja a *Measure of Agreement Kappa* értékét, ami esetünkben κ=.63. Ebből arra következtetünk, hogy a κ együttható értéke jó, a bizottsági tagok közötti nézetazonosság tapasztalható.

A PSPP egyik korlátja, hogy a kappa teszthez nem jeleníti meg a szignifikancia szintet (az *Approx. Sig.* mező üresen maradt). Ugyanezzel az adatbázissal az SPSS $p<.01$ szignifikancia szintet mutatott.

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

A két bizottsági tag értékelései között nézetazonosság tapasztalható, a *Cohen-féle kappa együttható* statisztikailag szignifikáns és értéke jó egyezésre utal (κ =.63, p<.01).

Jelölése:

$$\boldsymbol{\kappa = .63,\ p < .01}$$

ahol kappa együttható = a teszt eredménye, *p* érték

12.5.6. Mann-Whitney U teszt

A *Mann-Whitney U teszt* során két független csoport közötti különbséget vizsgálhatunk egy folytonos változóhoz viszonyítva. Mediánok összehasonlításán alapszik. Paraméteres megfelelője a független mintás t-próba. A folytonos változó pontszámait rangsorolja a két csoportra vetítve. A rangokból következtethetünk a csoportok közötti különbség irányára.

Feltételei:

– egy kategorikus változó, ami 2 csoportot jelöl (pl. nemek);

– egy folytonos változó.

A teszt a folytonos változó helyett ordinális változóval is elvégezhető, mivel lehetséges a mediánok kiszámítása.

Parancssor és beállítások:

SPSS: **Analyze → Nonparametric Tests → Legacy Dialog → 2 Independent Samples**

PSPP: A program grafikus felületéről nem érhető el a Mann-Whitney U teszt[30], viszont a *Syntax editor*-ban beprogramozható.

File→New→Syntax

Kétsoros parancssort írunk be:

NPAR TESTS

/MANN-WHITNEY = *folytonosváltozó* BY *csoportosváltozó* (1,2).

A "*folytonosváltozó*" helyett az adatbázisunkban használt változó nevét adjuk meg.

A "*csoportosváltozó*" helyett az adatbázisunkban szereplő csoportosító változónkat adjuk meg, ahol a két csoportot 1-el és 2-vel kódoltuk.

30 A teszt beállításairól bővebben lásd a PSPP kezelési útmutató 15.11.8-as szekcióját. A PSPP kezelési útmutatója a következő linken érhető el: https://www.gnu.org/software/pspp/manual/pspp.pdf

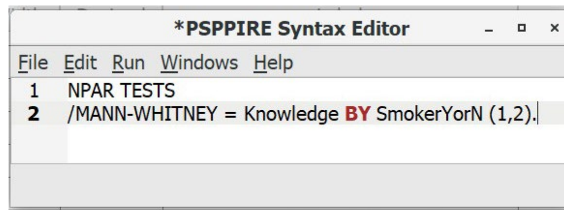
Majd a *Run* menüpontból lefuttatjuk az *All* menüpontra kattintva. Az eredményeket a kimeneti ablakban találjuk.

A vizsgálati hipotézis:
H_0 – a csoportok között nincs különbség.
H_1 – a csoportok között különbség van.

Az eredmény értelmezése:
Ha a teszt eredménye nem szignifikáns ($p > .05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a csoportok nem különböznek egymástól.

Ha a teszt eredménye szignifikáns ($p < .05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a csoportok között különbség van. Ebben az esetben szükségünk lesz a medián értékekre is!

Magyarázat egy példán keresztül:
A példánk szerint a két független csoportot a dohányzó (n = 117) és a nem dohányzó (n=198) alanyok alkotják (n = 315). A folytonos változónk azt méri egy skálán[31], hogy mennyire vannak tisztában az alanyok a 2016-ban életbe léptetett korlátozásokkal.



A *Syntax Editor*-ba beírjuk a fent megadott parancssort, az adatbázisban szereplő változók neve szerint (a tudásszint folytonos változója a *Knowledge*, a csoportosító változónk *SmokerYorN*, dichotóm értékei *1 - nem* és *2 - igen*).

```
NPAR TESTS

NPAR TESTS
/MANN-WHITNEY = Knowledge BY SmokerYorN (1,2).
```

Ranks

	N			Mean Rank		Sum of Ranks	
	No	Yes	Total	No	Yes	No	Yes
Knowledge about restrictions	117,00	198,00	315,00	138,29	169,65	16179,50	33590,50

Test Statistics

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
Knowledge about restrictions	9276,50	16179,50	-3,04	,002

31 Jelen példa szempontjából a folytonos változó kialakítása mellékszál, viszont elmondható, hogy a konkrét kutatásban több ismeretet mértünk, majd az egyes kérdésekre adott helyes válaszokat pontoztuk és a pontokat összeadva alakítottuk ki az ismeretekre vonatkozó, a példánkban is szereplő változót.

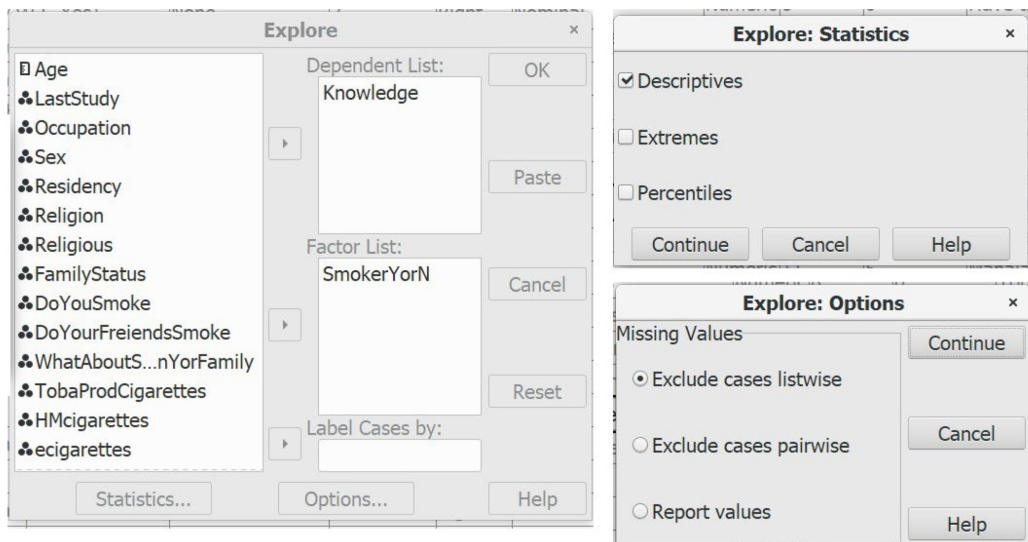
A teszt eredménye szignifikáns (p = .002), tehát elvetjük a nullhipotézist és megállapítjuk, hogy a csoportok között különbség van. Azt, hogy melyik csoport tájékozottabb a folytonos változónk szerint, a *Mean Rank* értékei alapján tudjuk megállapítani. Mivel a dohányosok csoportjának válaszai szerinti rangok átlaga nagyobb, ezért őket tekinthetjük tájékozottabbaknak.

Viszont, az eredmények megjelenítésekor még egy művelet hátra van: ki kell számoljuk a csoportok medián értékeit a folytonos változóra nézve.

Ez a művelet SPSS-ben a következő:
SPSS: **Analyze → Compare Means → Means**

Az *Independent list*-be a csoportosító változónkat tesszük, a *Dependent list*-be a folytonos változónkat, majd az *Options* fül alatt hozzáadjuk a *Cell Statistics* ablakhoz a *Median* műveletet.

PSPP-ben az **Analyze → Descriptive statistics → Explore** parancssort követjük az alábbi beállításokkal:



A *Factor List*-be tesszük a csoportosító változónkat, a *Dependent List*-be a folytonos változónkat. A művelet lefuttatása után összesen négy táblázatot kapunk. Ezek közül minket a két utolsó táblázat, a *Case Processing Summary* és a *Descriptives* fog érdekelni, ugyanis a csoportonkénti esetszámok és a medián értékek ezekből olvashatók ki:

Case Processing Summary

	Do you smoke? Yes or No	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Knowledge about restrictions	No	117	100%	0	0%	117	100%
	Yes	198	100%	0	0%	198	100%

Descriptives

	Do you smoke? Yes or No			Statistic	Std. Error
Knowledge about restrictions	No	Mean		7,91	,16
		95% Confidence Interval for Mean	Lower Bound	7,59	
			Upper Bound	8,23	
		5% Trimmed Mean		8,00	
		Median		8,00	
		Variance		3,06	
		Std. Deviation		1,75	
		Minimum		1,00	
		Maximum		11,00	
		Range		10,00	
		Interquartile Range		2,00	
		Skewness		-,96	,22
		Kurtosis		1,93	,44
	Yes	Mean		8,53	,10
		95% Confidence Interval for Mean	Lower Bound	8,33	
			Upper Bound	8,72	
		5% Trimmed Mean		8,56	
		Median		9,00	
		Variance		1,89	
		Std. Deviation		1,37	
		Minimum		3,00	
		Maximum		12,00	
		Range		9,00	
		Interquartile Range		1,00	
		Skewness		-,44	,17
		Kurtosis		,91	,34

Azok, akik nem dohányoznak, a folytonos változó tekintetében 8-as mediánt, akik dohányoznak, 9-es mediánt értek el.

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

Az eredmény bemutatásakor meg kell adni a mindkét csoportra vonatkozó mediánt és az eseti számot, a teszt ereményét, a *z* pontszámot és a szignifikancia szintet (SPSS-ben a korrelációs együtthatót is, amit a PSPP nem számol).

A *Mann-Whitney U Teszt* szerint szignifikáns különbség van a nem dohányzók ($Md = 8$, $n = 117$) és a dohányzók csoportja ($Md = 9$, $n = 198$) között a megszorításokkal kapcsolatos intézkedések ismeretének tekintetében ($U = 9276{,}5$, $z = –3.04$, $p<.01$). Megállapíthatjuk, hogy a dohányzó alanyok magasabb pontszámot értek el a restrikciók ismeretét illetően, mint a nem dohántyzó alanyok. Ennek lehetséges magyarázata…. (pl. őket érinti jobban)[32].

32 A kutatás eredményeinek lehetséges magyarázatai lehetőleg magából a kutatásból származzanak. Ha a kérdőívekből erre nem derül fény, akkor ideális esetben kiegészítjük a kutatást, akár minőségi módszerekkel (interjú, fókuszcsoport). További lehetőség, hogy a szakirodalomban felmerülő tanulmányok megfigyelései és konklúziói alapján fogalmazzuk meg a lehetséges magyarázatokat.

Jelölése:

Md = 8, n = 117

ahol *Md* a median és *n* a csoport eseti száma

U = 9276,5, z = –3.04, $p < .01$, r = .06.[33]

ahol *U* = a teszt eredménye, *z* egy statisztikai mérőszám[34],
p érték a szignifikancia szint és *r* a korrelációs együttható

12.5.7. Wilcoxon féle előjeles rang teszt

A *Wilcoxon féle előjeles rang teszt* olyankor alkalmazható, amikor az alanyokat különböző időpontokban vagy különböző körülmények között ismételten mérjük fel. A teszt a pontszámokat rangokká alakítja, majd összeveti őket a két vizsgált időpontban, vagy kritérium szerint. Mivel nem parametrikus tesz, ebben az esetben is a próba mediánok összevetésén alapszik. A teszt arra ad választ, hogy az alanyok pontszámai változtak-e a két időpont vagy két körülmény függvényében. Parametrikus megfelelője a páros mintás t-teszt.

Feltételei:

– két folytonos változó, melyek két különböző időpontban vagy körülmény kapcsán felvett pontszámokat tartalmaznak;

– a folytonos változókat képező skálák legyenek azonosak;

– mindkét változó ugyanattól a mintától, csoporttól származzon.

A teszt ordinális változókkal is elvégezhető, mivel lehetséges a mediánok kiszámítása.

Parancssor és beállítások:

SPSS: **Analyze → Nonparametric Tests → Legacy Dialog → 2 Related Samples**

A két folytonos változót a *Test pairs* ablakba tesszük. A *Test Type*-nál bejelöljük a *Wilcoxon* tesztet.

Az *Options* fül alatt kiválasztjuk a *Quartiles* bállítást.

PSPP: **Analyze → Nonparametric Tests →2 Related Samples Test**

A vizsgálati hipotézis:

H_0 – a folytonos változók pontszámai azonosak.

H_1 – a folytonos változók pontszámai között különbség van.

33 Az *r* együttható értéke a művelet kapcsán csak az SPSS-ben elérhető, a PSPP nem adja meg.

34 Leírja egy érték viszonyát egy értékcsoport átlagához képest.

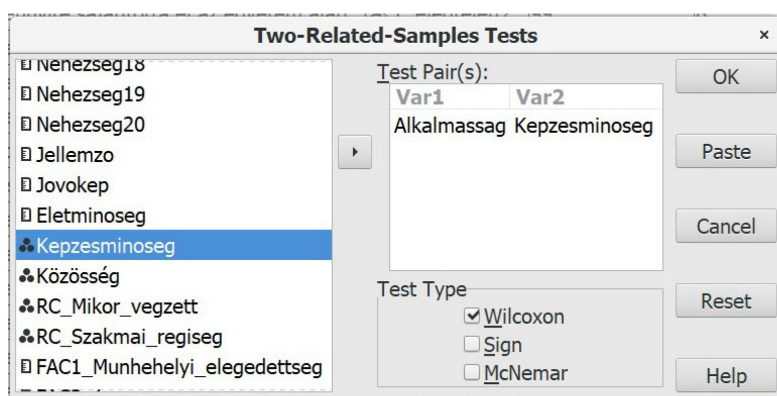
Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns ($p > .05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a folytonos változók pontszámai azonosak.

Ha a teszt eredménye szignifikáns ($p < .05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a folytonos változók pontszámai között különbség van. Ebben az esetben szükségünk lesz a medián értékekre is!

Magyarázat egy példán keresztül:

Példaként egy 196 alannyal végzett képzéshasznosulási kutatást hoznék fel. Az alanyok két azonos skálán értékelték a szakmai alkalmasságukat és az egyetemi képzés minőségét. Azt akartuk megvizsgálni, hogy a pontszámok között van-e különbség, vagyis ebben a példában nem két időpontban végzett mérést, hanem két körülményre vonatkozó mérést teszteltünk. Ezt a két változót (az adatbázisban az elnevezésük „*Alkalmassag*" és „*Kepzesminoseg*") párba állítjuk a *Test Pair(s)* ablakba, majd a *Test Type* opciói közül kiválasztjuk a *Wilcoxont*.



A teszt lefuttatása után a kimeneti ablakban a következő eredményt kaptuk:

```
NPAR TEST
    /WILCOXON Alkalmassag WITH  Kepzesminoseg (PAIRED).
```

Ranks

		N	Mean Rank	Sum of Ranks
Hogyan értékeli szakmai alkalmasságát? - Mennyire elégedett az egyetemi képzéssel?	Negative Ranks	21	43,14	906,00
	Positive Ranks	102	65,88	6720,00
	Ties	73		
	Total	196		

Test Statistics

	Hogyan értékeli szakmai alkalmasságát? - Mennyire elégedett az egyetemi képzéssel?
Z	-7,57
Asymp. Sig. (2-tailed)	,000

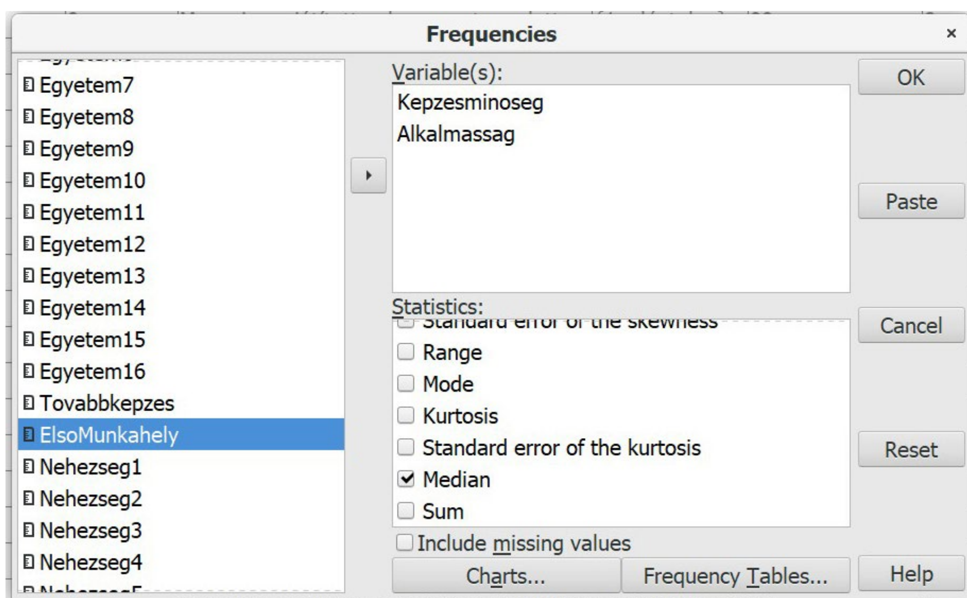
Mivel a *Wilcoxon teszt* eredménye szignifikáns ($p < .01$), elvetjük a nullhipotézist és arra következtetünk, hogy a folytonos változóink pontszámai között különbség van. Ebben az esetben a medián értékeket is külön ki kell számítanunk, amit az alábbi módon tehetünk meg:

SPSS: Analyze → Descriptive Statistics → Frequencies

A *Variables* ablakba betesszük mindkét változónkat. Az *Statistics* gombbal előhívjuk a további beállításokat és a *Central tendency*-ben bejelöljük a *Median* műveletet.

PSPP: Analyze → Descriptive statistics → Frequencies

A *Variables* ablakba betesszük mindkét változónkat. Az *Statistics* ablakban bejelöljük a *Median* műveletet.



A teszt lefuttatása után több táblázatot is kapunk, minket a következő két táblázat fog érdekelni:

Hogyan értékeli szakmai alkalmasságát?

N	*Valid*	196
	Missing	38
Mean		3,96
Percentiles	50 (Median)	4

Mennyire elégedett az egyetemi képzéssel?

N	*Valid*	196
	Missing	38
Mean		3,22
Percentiles	50 (Median)	3

A szakmai alkalmasság pontszámának mediánja $Md = 4$, az egyetemi képzéssel való elégedettségé $Md = 3$.

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

A *Wilcoxon féle előjeles rang-teszt* statisztikailag szignifikáns különbséget talált a szakmai alkalmasság ($Md = 4$, $n = 196$) és az egyetemi képzéssel való elégedettség ($Md = 3$, $n = 196$) között ($z = -7.57$, $p < .01$).

Jelölése:

Md = 3, n = 196

ahol *Md* a median és *n* az alanyok eseti száma

z = –7.57, p < .01

ahol *z* egy statisztikai mérőszám[35], *p* érték a szignifikancia szint

12.5.8. Kruskall-Wallis teszt

A *Kruskall-Wallis teszt* során három vagy több csoport pontszámai hasonlíthatóak össze egy folytonos változón. Hasonlít a *Mann-Whitney U teszt*re, csak ezzel a próbával nem kettő, hanem három vagy több független csoport közötti különbséget vizsgálhatjuk. Mediánok összehasonlításán alapszik. Parametrikus megfelelője az egyirányú csoporton belüli ANOVA. Gyakran használják például az életkor-kategóriák szerinti mérések összehasonlítására.

Feltételei:

– egy kategorikus független változó három vagy több kategóriával;
– egy folytonos függő változó;
– a csoportokban más és más alanyok vesznek részt.

A teszt ordinális változókkal is elvégezhető, mivel lehetséges a mediánok kiszámítása.

Parancssor és beállítások:

SPSS: Analyze → Nonparametric Tests → Legacy Dialog → K Independent

A *Test Variable List* ablakba betesszük a folyamatos változónkat, ez egyben a függő változónk. A *Grouping Variable* ablakba betesszük a három vagy több kategóriát tartalmazó csoportokat jelölő független változónkat. A *Define Range* fül alatt megadjuk a kategorikus változónk csoportjait jelző minimum és maximum értékeket. Vagyis ha három csoportunk van és 1-től 3-ig kódoltuk őket, akkor a *Minimum* 1, a *Maximum* 3 lesz. A *Test Type* beállításoknál bejelöljük a *Kruskal-Wallis H*-t.

PSPP: Analyze → Nonparametric Statistics → K Independent Samples

A *Test Variable List* ablakba betesszük a folyamatos változónkat, ez egyben a függő változónk.

A *Grouping Variable* ablakba betesszük a három vagy több kategóriát tartalmazó csoportokat jelölő független változónkat. A *Define Groups* fül alatt megadjuk a kategorikus változónk csoportjait jelző minimum és maximum értékeket. Vagyis ha három csoportunk van és 1-től 3-ig kódoltuk őket, akkor a *Lower limit* 1, a *Upper limit* 3 lesz. A *Test Type* beállításoknál bejelöljük a *Kruskal-Wallis H*-t és a *Mediánt*.

35 Leírja egy érték viszonyát egy értékcsoport átlagához képest.

A vizsgálati hipotézis:

H_0 – a csoportok medián értékei egyenlőek.

H_1 – a csoportok medián értékei különböznek.

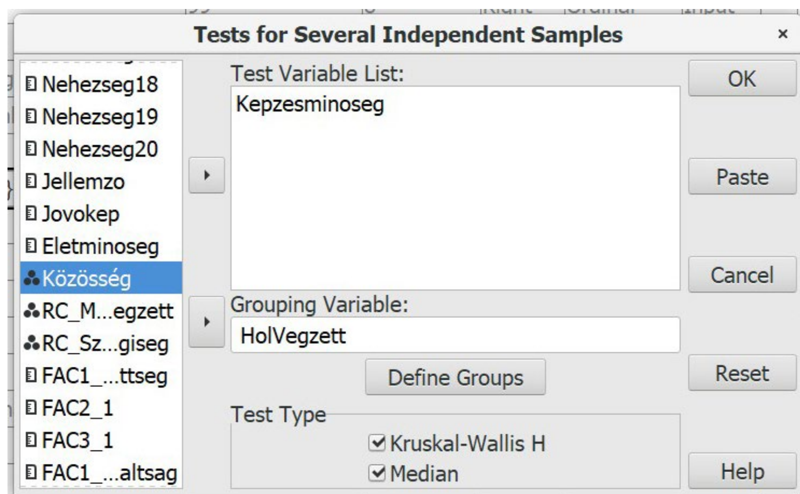
Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns ($p > .05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a csoportok mediánjai nem különböznek egymástól.

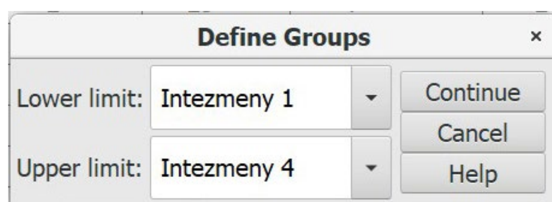
Ha a teszt eredménye szignifikáns ($p < .05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a csoportok mediánjai különböznek. Ebben az esetben szükségünk lesz a medián értékekre is.

Magyarázat egy példán keresztül:

Példánk alapjául egy képzéshasznosulással kapcsolatos adatbázis szolgált, melyet 196 alany töltött ki. A kategorikus változónk 4 kategóriát tartalmaz, mely egy-egy oktatási intézményt jelöl (ez lesz a *Grouping Variable*). A folytonos változónk a képzés minőségével való elégedettséget fejezi ki egy 1-től 5-ig terjedő skálán (ez lesz a *Test Variable List*-ben a „*Kepzesminoseg*" nevű változónk).



Mivel a kategorikus változónk 4 csoportot tartalmaz, melyek numerikus kódolása 1-től 4-ig tart, a *Define Groups* opciót a következőképpen állítjuk be:



A teszt lefuttatása után a PSPP kimeneti ablakában a következő eredményeket kaptuk:

```
NPAR TEST
     /KRUSKAL-WALLIS = Kepzesminoseg BY HolVegzett (1, 4)
     /MEDIAN = Kepzesminoseg BY HolVegzett (1, 4).
```

Ranks

	Hol végzett?	N	Mean Rank
Mennyire elégedett az egyetemi képzéssel?	Intezmeny 1	46	94,50
	Intezmeny 2	83	104,19
	Intezmeny 3	49	97,38
	Intezmeny 4	18	85,56
	Total	196	

Test Statistics

	Mennyire elégedett az egyetemi képzéssel?
Chi-Square	2,28
df	3
Asymp. Sig.	,516

Frequencies

		Hol végzett?			
		Intezmeny 1	*Intezmeny 2*	*Intezmeny 3*	*Intezmeny 4*
Mennyire elégedett az egyetemi képzéssel?	*> Median*	18	37	20	6
	≤ Median	28	46	29	12

Test Statistics

	N	*Median*	*Chi-Square*	*df*	*Asymp. Sig.*
Mennyire elégedett az egyetemi képzéssel?	196	3,00	,93	3	,817

A teszt eredménye szempontjából a bekarikázott részeket értelmezzük: A teszt *Statistics* táblázatból kitűnik, hogy az eredmény statisztikailag nem szignifikáns ($p = .516$, vagyis $p > .05$), tehát fenntartjuk a nullhipotézist, mely szerint a csoportok medián értéke ($Md = 3$) nem különbözik. Az eredmények megjelenítésekor raportáljuk a khi-négyzet értékét és a szabadsági fokokat is ($X^2 = 2.28$, $df = 3$). Jelen esetben nincs több dolgunk, mert a teszt nem volt szignifikáns.

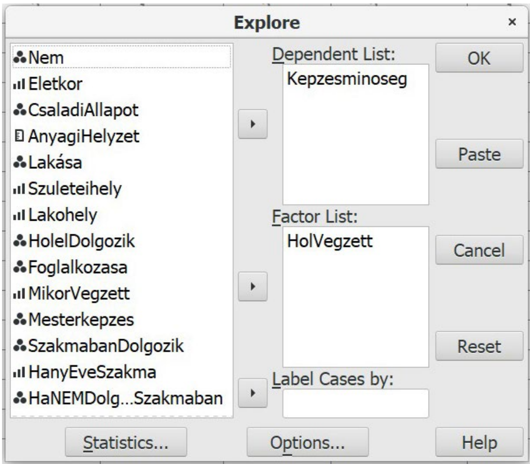
Ha a teszt eredménye szignifikáns lett volna ($p < .05$), akkor megvizsgálnánk, hogy a 4 intézmény közül melyek kaptak magasabb pontszámokat. Erre utal a *Ranks* táblázat; a *Mean Rank* tükrözné az intézmények által nyújtott képzések közötti eltéréseket, viszont ez csak számunkra jelenthet tájékozódási pontot, a dolgozatba a medián értékeket kellene leírjuk. Ennek magyarázata az adatok nemnormális eloszlásában rejlik. Ilyen esetekben a medián érték pontosabban írja le az adatokat, mint az átlag. Átlagot normális eloszlású adatoknál használhatnánk.

Ha a teszt eredménye szignifikáns lett volna (de jelen példában nem az!), szükségünk lenne a képzésminőség 4 intézményre vetített mediánjára (vagyis a folytonos változónak a kategorikus változók szerinti medián értékeire). Ezt a *Mann-Whitney U teszt*nél már ismertetett módon kaphatjuk meg.

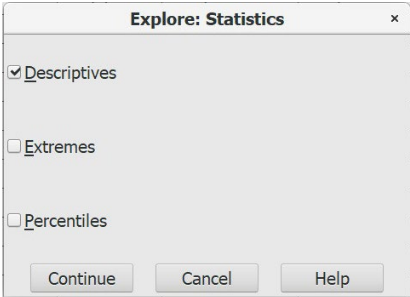
SPSS: Analyze → Compare Means → Means

Az *Independent list*-be a csoportosító változónkat tesszük, a *Dependent list*-be a folytonos változónkat, majd az *Options* fül alatt hozzáadjuk a *Cell Statistics* ablakhoz a *Median* műveletet.

PSPP-ben az **Analyze → Descriptive statistics → Explore** parancssort követjük az alábbi beállításokkal: A *Factor List*-be tesszük a csoportosító változónkat, a *Dependent List*-be a folytonos változónkat.



A *Statistics* fül alatt beállítjuk a *Descriptives* opciót, a többi beállítás az alapértelmezett marad.



A művelet lefuttatása után összesen négy táblázatot kapunk. Ezek közül minket a két utolsó táblázat, a *Case Processing Summary* és a *Descriptives* fog érdekelni, ugyanis a csoportonkénti esetszámok és a medián értékek ezekből olvashatók ki. A példa kedvéért csak egy táblázatrészt közlünk. Ezeket az adatokat az utolsó táblázat megadja a kategorikus változónk összes kategóriájára, a medián értéket az aláhúzott sorban találjuk.

Hol végzett?			*Statistic*	*Std. Error*
Intezmeny 1	Mean		3,11	,17
	95% Confidence Interval for Mean	Lower Bound	2,76	
		Upper Bound	3,46	
	5% Trimmed Mean		3,12	
	Median		3,00	
	Variance		1,39	
	Std. Deviation		1,18	
	Minimum		1,00	
	Maximum		5,00	
	Range		4,00	
	Interquartile Range		1,00	
	Skewness		-,56	,35
	Kurtosis		-,31	,69

Jelen példában mind a négy intézmény pontszámainak medián értéke *Md* = 3 volt, az a tény, hogy az átlagok (*Mean* vagy *M*) különböztek, nem mérvadó, mivel az adataink nem-normális eloszlásúak voltak.

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

A *Kruskal-Wallis Test* eredménye statisztikailag nem szignifikáns a kutatásban szereplő négy intézmény képzésminőségének megítélése szempontjából (Intézmény1: *n* = 46, *Md* = 3; Intézmény2: *n* = 83, *Md* = 3; Intézmény3: *n* = 49, *Md* = 3; Intézmény 4: *n* = 18, *Md* = 3), $\chi 2$ (3, n = 196) = 2.28, p > .05.

Jelölése:

***n* = 46, *Md* = 3**

ahol *n* az eseti szám, *Md* a medián

***χ2* (3, *n* = 196) = 2.28, *p* > .05.**

ahol khi-négyzet (*df* szabadsági fokok, eseti szám) = a teszt eredménye, *p* szignifikancia szint

12.5.9. Friedman teszt

A *Friedman próba* ugyanazt a mintát három vagy ennél több alkalommal, illetve kondíció mellett hasonlítja össze. Az alanyok nincsenek csoportokra osztva. Rangok átlagának[36] összehasonlításán alapszik. Paraméteres megfelelője az egyirányú ismételt méréses varianciaanalízis (ANOVA).

A Freiedman teszt hasonlít a Wilcoxon féle előjeles rang teszthez, mivel mindkét teszt rangsoroláson alapuló nem parametrikus teszt, de a Friedman teszt kevésbé érzékeny. A Friedman teszt csak rangsorolja az eseteket, míg a Wilcoxon teszt figyelembe veszi az esetek közötti különbségek magnitúdóját.

Feltételei:

– egy minta (ugyanazok az alanyok);

– három vagy több folytonos vagy ordinális változó, melyek azonosak;

– a változóknak megfelelő három vagy több időpontban történt mérés vagy körülmény.

Parancssor és beállítások:

SPSS: Analyze → Nonparametric Tests → Legacy Dialog → K Related Samples

A *Test Type* opciói közül kiválasztjuk a *Friedman* tesztet. A *Statistics* fül alatt kiválasztjuk a *Quartiles* bállítást.

36 A rangok átlaga: a megfigyelés elemeit rangsorba állítjuk, majd kiszámítjuk a rangok átlagát.

PSPP: Analyze → Non-Parametric Statistics → K Related Samples

A *Test Type* opciói közül kiválasztjuk a *Friedman* tesztet (ez az alapértelmezett teszt).

A vizsgálati hipotézis:

H_0 – a mérések rangjainak átlagai megegyeznek.
H_1 – a mérések rangjainak átlagai között különbség van.

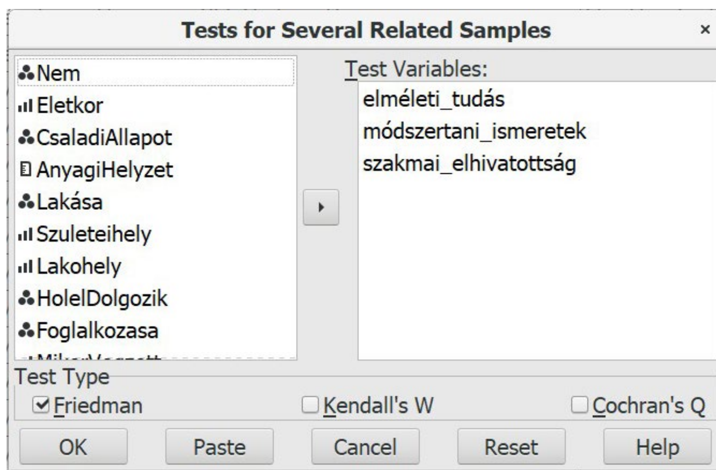
Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns ($p > .05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a mérések rangjainak átlaga között nincs különbség.

Ha a teszt eredménye szignifikáns ($p < .05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a mérések rangjainak átlaga között különbség van. Ebben az esetben ki kell számolnunk a mediánokat is.

Magyarázat egy példán keresztül:

A példánk alapjául egy olyan kutatás szolgál, amely során egy egyetemi képzés nyújtotta ismeretek és készségek szintjét mértük fel: az elméleti tudás elsajátítását, a módszertani ismeretek elsajátítását és a szakmai elhivatottság kialakulását. Mindhárom változót 1-től 5-ig terjedő skála alapján mértük.



A teszt lefuttatása után a kimeneti ablakban a következő eredményeket kaptuk:

```
NPAR TEST
     /FRIEDMAN = elméleti_tudás módszertani_ismeretek szakmai_elhivatottság.
```

Ranks

	Mean Rank
Mennyire sajátította el az egyetem alatt: elméleti tudás	2,09
Mennyire sajátította el az egyetem alatt: módszertan	1,79
Mennyire sajátította el az egyetem alatt: elhivatottság	2,12

Test Statistics

N	189
Chi-Square	26,69
df	2
Asymp. Sig.	,000

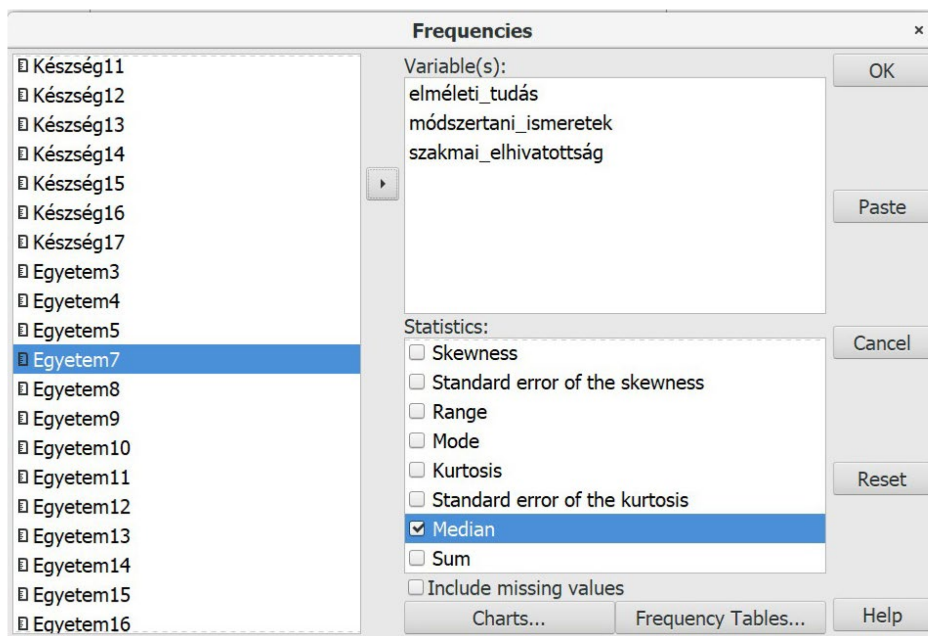
A *Friedman teszt* eredménye szignifikáns ($p < .01$), tehát elvetjük a nullhipotézist és megállapítjuk, hogy a mérések között statisztikailag szignifikáns különbség van. A *Mean Rank* adataiból, vagyis a rangok átlagából következtethetünk a változók közötti különbségek nagyságára. Sőt, a medián értékeket is megadhatjuk, amelyet az alábbi módon számolhatunk ki:

SPSS: Analyze → Descriptive Statistics → Frequencies

A *Variables* ablakba betesszük mindkét változónkat. A *Statistics* gombbal előhívjuk a további beállításokat és a *Central tendency*-ben bejelöljük a *Median* műveletet.

PSPP: Analyze → Descriptive statistics → Frequencies

A *Variables* ablakba betesszük mindkét változónkat. A *Statistics* ablakban bejelöljük a *Median* műveletet.



A kimeneti ablakot összegezve, mindhárom változónk medián értéke *Md* = 4. Ebben a konkrét példában ez az adat nem segít az értelmezésben, ezért a rangok átlagára hagyatkozunk.

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**
A *Friedman teszt* eredménye a három vizsgált készségre (elméleti tudás, módszertan, elhivatottság) statisztikailag szignifikáns $\chi 2$ (2, n = 189) = 26,69, p<.01), azaz a készségek rangjainak átlaga eltérő. A legmagasabb értéket a szakmai elhivatottság kialakulása érte el (*MR* = 2,12), ezt követi az elméleti tudás elsajátítása (*MR* = 2,09), majd a legkisebb értéket a módszertani ismeretek elsajátítása (*MR* = 1,79) mutastta.[37]

Jelölése:

(MR = 1,79)
ahol *MR* a rangok átlaga

$\chi 2$ (2, n = 189) = 26,69, p < .01),
ahol khi-négyzet (*df* szabadsádi fokok, eseti szám) = a teszt eredménye,
p szignifikancia szint

12.6. Paraméteres tesztek

12.6.1. Egymintás t-teszt

Az egymintás t-próbával megállapíthatjuk, hogy egy folytonos függő változó átlaga hogyan viszonyul egy előre meghatározott referencia értékhez. Mivel parametrikus próba, ezért a folytonos változó normális eloszlású kell, hogy legyen. Szükséges a folytonos változó előzetes normalitásvizsgálata, melyet a *Kolmogorov-Smirnov* vagy a *Saphiro-Wilk tesztek* egyikével végezhetünk el a 11.2.9. alfejezet szerint. A gyakorlatban például olyankor alkalmazhatjuk, amikor egy saját adatgyűjtésű változónkat szeretnénk egy előzetesen ismert kutatási eredményhez viszonyítani. Értelemszerűen akkor kaphatunk megbízható eredményt, ha a saját változónk kódolása, intervalluma megegyezik a referencia kutatásban használtéval. Ennek tipikus esete lehet, amikor a kérdőívünket más kutatások kérdőívei inspirálták (ugyanazt a mérőeszközt használtuk), vagy amikor egy saját, már kiértékelt kutatást ismételünk meg.

37 A rangok átlaga (*MR*) helyett a medián (*M*) értékeket is írhatjuk.

Feltételei:

– egy intervallum mérési szintű változó;

– egy előre meghatározott referencia érték, amihez hasonlítani akarjuk a változónk átlagát;

– a változó adateloszlása megfelel a normalitás kritériumainak.

Parancssor és beállítások:

Az **SPSS** és **PSPP** parancssora megegyezik:

Analyze → Compare Means → One - Sample T Test

A *Test Variable* ablakba tesszük a függő intervallum típusú változót, a *Test Value* cellába beírjuk azt az értéket, amivel össze szeretnénk hasonlítani az intervallum átlagát.

A vizsgálati hipotézis:

H_0 – a függő változó átlaga megegyezik az előre meghatározott értékkel.

H_1 – a függő változó átlaga eltér az előre meghatározott értéktől.

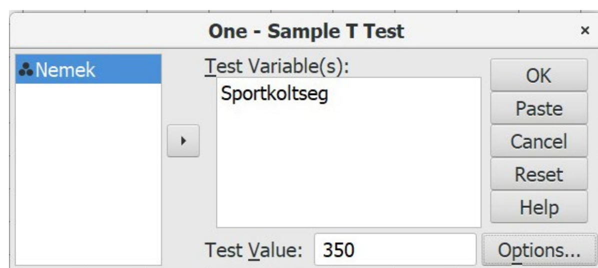
Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns ($p > .05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a változó átlaga megegyezik az előre meghatározott értékkel.

Ha a teszt eredménye szignifikáns ($p < .05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a változó átlaga eltér az előre meghatározott értéktől.

Magyarázat egy példán keresztül:

A példa kedvéért feltételezzük, hogy lekérdeztünk egy mintát ($n = 55$) a havi sporttevékenységekre szánt kiadásaikról. Ez képezi a folytonos, normális eloszlású változónkat. Az értékek egy tetszőleges pénznemben kifejezett havi költséget jelölnek. Szintén a példa kedvéért feltételezzük, hogy rendelkezésünkre állnak hivatalos statisztikai adatok, melyek szerint egy adott ország populációja átlagban 350 tetszőleges pénznemet költ sporttevékenységekre (ez lesz a referencia értékünk).[38]



38 Didaktikai célzatból készült fiktív adatbázisról és referencia értékről van szó. A folytonos változónkat úgy generáltuk, hogy normál adateloszlású legyen, mely feltétele a tesztnek. A használt program linkje: https://www.statology.org/normal-distribution-dataset-generator/

A fentiek szerint beállítottuk és lefuttattuk a műveletet, majd a kimeneti ablakban az alábbi eredményt kaptuk:

```
T-TEST /TESTVAL=350
    /VARIABLES= Sportkoltseg      /MISSING=ANALYSIS
    /CRITERIA=CI(0.95).
```

One-Sample Statistics

	N	*Mean*	*Std. Deviation*	*S.E. Mean*
Összeg	55	508,67	201,25	27,14

One-Sample Test

	Test Value = 350,000000					
					95% Confidence Interval of the Difference	
	t	*df*	*Sig. (2-tailed)*	*Mean Difference*	*Lower*	*Upper*
Összeg	5,85	54	,000	158,67	104,27	213,08

Azt látjuk, hogy az egymintás t-teszt eredménye szignifikáns ($p < .000$, ezt a dolgozatban úgy jelöljük, hogy $p < .01$). Az eredmény értelmezése a következő: mivel a teszt statisztikailag szignifikáns, elvetjük a nullhipotézist és megállapítjuk, hogy a változó átlaga eltér az előre meghatározott értéktől. A teszt konkrétan a folytonos változónk átlagát is megadja, ami jelen esetben az *One-Sample Statistics* táblázatból olvasható ki, ez lesz az átlag érték (*Mean*) $M = 508{,}67$. Mivel a referencia értékünk 350 volt, kijelenthetjük, hogy az általunk visgált populáció többet költ havi sporttevékenységekre, mint a hivatalos statisztikák szerinti országos átlag.

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

Egymintás t-teszttel vizsgáltuk a mintánk ($n = 55$) havi szinten sporttevékenységekre szánt kiadásait a statisztikai hivatal által közölt országos átlaghoz képest (M = 350). A teszt eredménye szignifikáns különbséget jelzett a mintánk átlaga ($M = 508{,}67$) és az országos átlag között, $t\ (54) = 5.85$, $p < .01$.

Jelölése:

$$\boldsymbol{n = 55; M = 350}$$

ahol *n* az eseti szám; *M* az átlag

$$\boldsymbol{t\ (54) = 5.85,\ p < .01}$$

ahol t-teszt (szabadsági fokok - *df*) = a teszt eredménye, *p* érték

12.6.2. Független mintás t-teszt

A független mintás t-próba során két különböző csoport (minta) átlagpontszámait hasonlíthatjuk össze kettő vagy több folytonos változóra nézve. Mivel parametrikus próba, ezért a folytonos változóknak normális eloszlásúnak kell lenniük. A teszt átlagpontszámok valószínűségi összehasonlításán alapszik. Amennyiben társadalomtudományok területén végzünk adatgyűjtés, az emberek véleményei, attitűdjei, ismeretei és egyéb attribútumai csak a legritkább esetekben fognak normális eloszlású folytonos változót képezni. A természeti jelenségekkel kapcsolatos megfigyeléseink és méréseink eredményezhetnek szimmetrikus, normális eloszlást, mint például a testmagasság, viszont ehhez viszonylag nagy elemszámú mintára lesz szükségünk. Nem-parametrikus megfelelője a *Mann-Whitney U Teszt.*

Feltételei:

– egy kategórikus dichotóm változó (mely két csoportot jelöl, pl. férfiak/nők), ez a független változónk;

– egy folytonos normális eloszlású változó (pl. havi sporttevékenységekre költött összeg), ez a függő változónk;

– a teszt alkalmazható kis elemeszámú mintán, de javasolt a minimum 30 eset.

Parancssor és beállítások:

SPSS: Analyze → Compare Means → Indepentent Samples T Test

A *Test variable* ablakba betesszük a folytonos változónkat, a *Grouping variable* ablakba a dichotóm (két csoportot jelölő) változónkat. A *Define groups* fül alatt megadjuk a csoportok kódolás szerinti értékeit (pl. *1* – férfi, *2* – nő).

PSPP: Analyze → Compare Means → Indepentent Samples T Test

A *Test variable(s)* ablakba betesszük a folytonos változónkat, a *Grouping variable* ablakba a dichotóm (két csoportot jelölő) változónkat. A *Define groups* fül alatt megadjuk a csoportok kódolás szerinti értékeit (pl. *1* – férfi, *2* – nő).

A vizsgálati hipotézis:

H_0 – a csoportok között a folytonos változón mért pontszámok átlaga nem különbözik.

H_1 – a csoportok között a folytonos változón mért pontszámok átlaga különbözik.

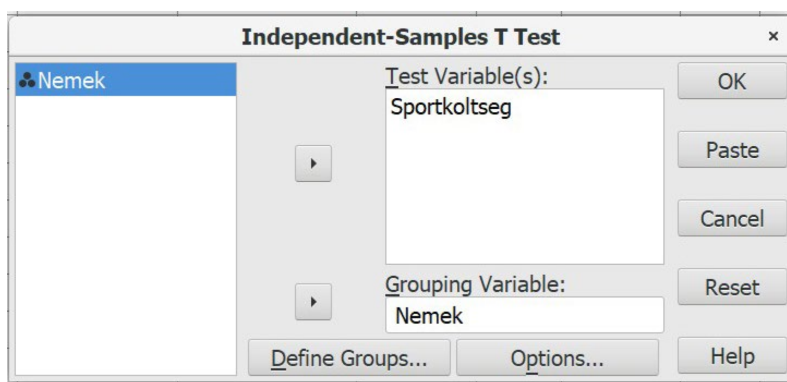
Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns ($p > .05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a független változók átlag pontszámai között nincs szignifikáns különbség (a két csoport között nincs különbség).

Ha a teszt eredménye szignifikáns ($p < .05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a független változók átlag pontszámai között szignifikáns különbség van (a két csoport között különbség van).

Magyarázat egy példán keresztül:

A példánkhoz egy fiktív adatbázist hoztunk létre, amlyben a független változónk 30 férfit és 25 nőt tartalmaz (ők képezik a két csoportot), a függő változónk a havi sporttevékenységekre költött összeget tartalmazza (a változónk röviden „Sportköltség"-nek neveztük). A függő változónk normális adateloszlású. Az adatfeldolgozás során az eloszlást a *Kolmogorov-Smirnov* vagy a *Saphiro-Wilk tesztek* egyikével ellenőrizhetjük (lásd a 10.2.9.-es alfejezetet).



Az *Options* fül alatt a beállításokatat hagyjuk az alapértelmezett értékeken. A teszt lefuttatása után a kimeneti ablakban a következő eredményeket kapjuk:

```
T-TEST /VARIABLES= Sportkoltseg
      /GROUPS=Nemek(1,2)     /MISSING=ANALYSIS
      /CRITERIA=CI(0.95).
```

Group Statistics

	Nemek	N	Mean	Std. Deviation	S.E. Mean
Összeg	Férfi	30	495,40	183,67	33,53
	Nő	25	524,60	223,34	44,67

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Összeg	Equal variances assumed	,27	,607	-,53	53,00	,597	-29,20	54,86	-139,24	80,84
	Equal variances not assumed			-,52	46,46	,604	-29,20	55,85	-141,60	83,20

Mivel a Levene teszt nem szignifikáns, $p > .05$, a táblázat első sorából nézzük az adatokat. Ha a Levene teszt szignifikáns lett volna, akkor a második sort néznénk.

Első lépés: a kimeneti ablakban először a *Levene-próba* varianciaegyenlőségi értéket nézzük. Amennyiben a szignifikancia szint értéke nagyobb, mint $p > .05$, egyenlő varianciára utal és a továbbiakban a táblázat elő sorában megjelenített eredményeket használjuk. Ha a *Levene*-próba szignifikancia szintje kisebb, mint $p < .05$, akkor ez arra utal, hogy a két csoport folytonos változóra vetített varianciája eltérő és a táblázat második sora szerint tájékozódunk.

Tehát a csoportok közötti kölönbségek vizsgálatára a t-teszt eredményeit használjuk a táblázat első vagy második sorából, annak megfelelően, hogy a *Levene-próba* szignifikáns volt-e vagy sem.

Második lépés: a t-teszt eredményének értelmezése. Mivel jelen példában a próba eredménye nem szignifikáns (p>.05, vagyis p=.597), megállapítjuk, hogy a független változó (vagyis a sporttevékenységekre szánt költség) átlag pontszámai a mintában szereplő férfiak és nők között nem különbözik szignifikánsan.

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

Független mintás t-próbával vizsgáltuk a férfiak és nők csoporja közti sporttevékenységekre szánt költségeket. A férfiak (*M* = 495.4, *SD* = 183.67) és nők (*M* = 524.6, *SD* = 223.34)[39] havi költsége statisztikailag nem különbözik szignifikánsan; *t* (53) = -.53, *p* > .05.

Jelölése:

M = 495.4, SD = 183.67

ahol *M* median, *SD* szórás

t (53) = -.53, p > .05.

ahol t-teszt (szabadsági fokok - *df*) = a teszt eredménye, *p* érték

12.6.3. Páros mintás t-teszt

A páros mintás t-tesz során ugyanannak a populációnak két folytonos változón mért válaszai hasonlíthatóak össze. A folytonos változók két különböző időpontban történt mérésre vagy két különböző körülményre vonatkoznak, ezeket nevezzük függő változónak. A próba feltétele, hogy folytonos változók normális eloszlásúnak legyenek és ennek megfelelően átlagpontszámok valószínűségi összehasonlításán alapszik. A független mintás t-teszt esetén tett megjegyzésünk itt is érvényes, mely szerint az adatgyűjtés során csak ritkán fog ez utóbbi feltétel teljesülni. Nem-parametrikus megfelelője a *Wilcoxon féle előjeles rang teszt*.

Feltételei:

– azonos populációtól származó minta;
– két folytonos, normális eloszlású függő változó;
– a teszt alkalmazható kis elemeszámú mintán, de javasolt minimum 30 eset.

39 Az *M* és *SD* értékeinek nagysága a változók kódolásától függ, akkor értelmezhető, ha megadjuk, hogy milyen skálán mértük a preferenciákat. Jelen esetben fiktív adatbázisról van szó és az értékeket didaktikai célzatból random generáltuk, úgy, hogy normál adateloszlású változókat kapjunk. A használt program linkje: https://www.statology.org/normal-distribution-dataset-generator/

Parancssor és beállítások:

SPSS és PSPP parancssora megegyezik: Analyze → Compare Means → Paired Samples T Test

A *Test pair(s)* ablakba betesszük a két folytonos változónkat, a többi beállítás marad az alapértelmezett. Lehetőségünk van egyszerre több változó párt is lefuttatni, de az eremények mindig páronként értelmezendőek. Ha páratlan számú változót szeretnénk összehasonlítani, ANOVA tesztet végzünk.

A vizsgálati hipotézis:

H_0 – a mért pontszámok átlaga nem különbözik.

H_1 – a mért pontszámok átlaga különbözik.

Az eredmény értelmezése:

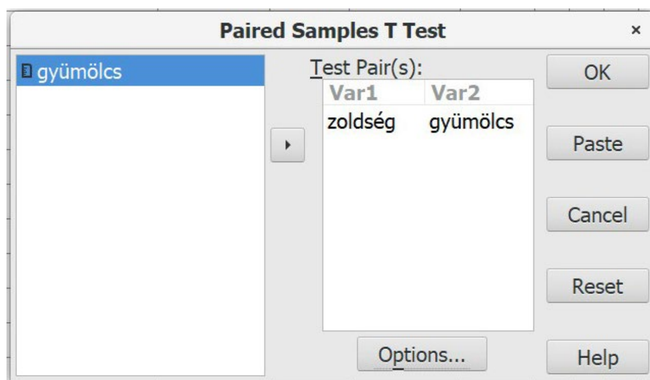
Ha a teszt eredménye nem szignifikáns ($p > .05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a függő változók átlag pontszámai között nincs szignifikáns különbség.

Ha a teszt eredménye szignifikáns ($p < .05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a függő változók átlag pontszámai között szignifikáns különbség van.

Magyarázat egy példán keresztül:

A képzeletbeli példánk egy 30-as eseti számú mintát (populációt) tartalmaz. Azt mértük, hogy az alanyok a zöldségeket és a gyümölcsöket mennyire kedvelik. Tehát jelen példában nem különböző időpontok eredményei, hanem két körülmény pontszámai képezik a folytonos, függő változónkat. Az adatbázis létrehozásakor – a példa kedvéért – normális eloszlású változókat hoztunk létre.

Megjegyzés: A kutatásunkban ne feledkezzünk meg az adatok előzetes normalitásvizsgálatáról, melyet a *Kolmogorov-Smirnov* vagy a *Saphiro-Wilk tesztek* egyikével végezhetünk el a 11.2.9. alfejezetben leírtak szerint.



A kimeneti ablakban a következő eredményt kaptuk:

```
T-TEST
    PAIRS = zoldség WITH  gyümölcs (PAIRED)
    /MISSING=ANALYSIS
    /CRITERIA=CI(0.95).
```

Paired Sample Statistics

		Mean	N	Std. Deviation	S.E. Mean
Pair 1	zoldség	284,77	30	74,72	13,64
	gyümölcs	303,33	30	89,05	16,26

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	zoldség & gyümölcs	30	-,01	,968

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	zoldség - gyümölcs	-18,57	116,68	21,30	-62,13	25,00	-,87	29	,391

A teszt eredményének értelmezése szempontjából a keretes részt nézzük: az eredmény nem szignifikáns (p = .391, vagyis p > .05), tehát fenntartjuk a nullhipotézist és megállapítjuk, hogy a függő változók átlag pontszámai nem különböznek statisztikailag szignifikánsan. Ebből az következik, hogy a minta alanyai egyformán kedvelik a zöldségeket és a gyümölcsöket.

Amennyiben a teszt eredménye szignifikáns lett volna (p < .05), akkor elvetnénk a nullhipotézist és megállapíthatnánk, hogy az alanyok preferenciái statisztikailag szignifikánsan különböznek (jelen példánkban ez nem áll fenn).

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

Páros mintás t-teszttel vizsgáltuk az alanyok (n=30) preferenciáit a zöldség- és a gyümölcsfogyasztás tekintetében. A két változó átlag pontszámai között nem tapasztaltunk szignifikáns különbséget (zöldség: M = 284.77, SD = 74.72; gyümölcs: M = 303.33, SD = 89.05)[40], t (29) = -.87, p > .05).

Jelölése:

***M* = 284.77, *SD* = 74.72**

ahol M median, SD szórás

***t* (29) = -.87, *p* > .05.**

ahol t-teszt (szabadsági fokok - *df*) = a teszt eredménye, p érték

40 Az M és SD értékeinek nagysága a változók kódolástól függ, akkor értelmezhető, ha megadjuk, hogy milyen skálán mértük a preferenciákat. Jelen esetben fiktív adatbázisról van szó és az értékeket didaktikai célzatból random generáltuk úgy, hogy normál adateloszlású változókat kapjunk. A használt program linkje: https://www.statology.org/normal-distribution-dataset-generator/

12.6.4. Egyutas (egytényezős) csoportok közötti ANOVA

Az egytényezős csoportok közötti ANOVA[41] több mint két csoport vagy kondíció átlag pontszámainak összehasonlítására alkalamas, egy folytonos változóra nézve. A folytonos változó normális adateloszlású kell legyen. Diszkrét és ordinális változók esetén az ANOVA próba nem alkalmazható, de a *Likert-skálák*[42] alkalmazása elfogadott, mivel ordinális jellegük ellenére a gyakorlatban metrikusnak tekinhetők.

A teszt egy *F* értéket ad meg, amely a varianciaelemzés során az eloszlás egyik értéke. A próba során arra kapunk választ, hogy a csoportok átlaga különbözik-e vagy sem. A csoportok átlagainak páronkénti viszonyulását ún. *post-hoc teszt*tel (utólagos vizsgálat) vizsgálhatjuk. Nem parametrikus megfelelője a *Kruskal-Wallis teszt*.

Feltételei:

– egy független változó több kategóriával (ezt nevezzük faktornak), amik jelölhetnek csoportokat vagy különböző kondíciókat;

– egy függő változó, mely folytonos és normális eloszlású.

Parancssor és beállítások:

Az **SPSS** és **PSPP** parancssora megegyezik, a beállítások hasonlóak:

Analyze → Compare means → One-way ANOVA.

A *Dependent List* ablakba tesszük a függő változónkat, a *Factor* ablakba a független változónkat (azt, amelyik csoportokat vagy kondíciókat tartalmaz).

Az **SPSS**-ben az *Options* fül alatt bejelöljük a következőket: *Descriptive, Homogeneity of variance test*. Az utólagos vizsgálathoz a *Post Hoc* fül alatt bejelöljük a *Tukey tesztet*.

A **PSPP**-ben a *Statistics* beállításoknál jelöljük be a következőket: *Descriptive, Homogeneity.*

A post-hoc teszthez a PSPP felhasználói felülete nem tartalmaz közvetlen beállítási lehetőséget, ezért a *Paste* fül alatt megnyítjuk a *Syntax Editor*-t és megadjuk a *Tukey teszt* parancsát:

ONEWAY /VARIABLES= *függő változónk neve* BY *független változónk neve* /POSTHOC=TUKEY

Majd a *Run* menüpontpól lefuttajuk az *All* parancsot.

---

41 ANOVA - variancia elemzést jelent.

42 Attitűdök mérésére használt skála, gyakori beosztása 1-től 5-ig, 7-ig vagy 10-ig terjed, ahol az „1"-es érték a leg negatívabb attitűdre, a maximális érték a leg pozitívabb attitűdre utal.

A vizsgálati hipotézis:
H_0 – a csoportok átlagai nem különböznek.
H_1 – a csoportok átlagai különböznek.

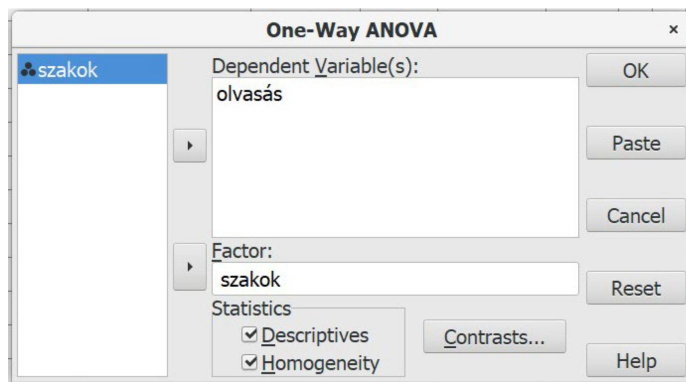
Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns ($p > .05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a csoportok átlagai nem különböznek.

Ha a teszt eredménye szignifikáns ($p < .05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a csoportok átlagai különböznek.

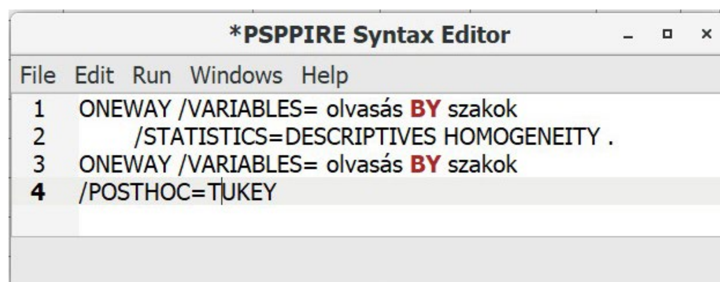
Magyarázat egy példán keresztül:

A példánkhoz létrehoztunk egy fiktív adatbázist, mely a reál, humán és művészeti szakos hallgatók (független változó, vagy faktor) olvasással töltött idejét tartalmazza (ez a folytonos változónk).[43] A beállítások az alábbi ablakban láthatók.



A post-hoc teszthez megnyítjuk a *Syntax Editort* a *Paste* gombra kattintva. Az első két sor adott az eddigi beállításaink alapján. Az első sort copy-paste-el duplázzuk és betesszük a 3. sorba. A 4. sorba beírjuk a már említett parancsot:

/POSTHOC=TUKEY



43 Didaktikai célzatból random generáltuk az adatbázist, a folytonos változónk normál adateloszlású. A használt program linkje: https://www.statology.org/normal-distribution-dataset-generator/

A *Run* menüpontból az *All*-ra kattintva lefuttatjuk a tesztet. A kimeneti ablakban a az alábbi eredmények jelennek meg, összesen öt táblázatban (a harmadik táblázat duplán fog szerepelni).

ONEWAY /VARIABLES= olvasás BY szakok
/STATISTICS=DESCRIPTIVES HOMOGENEITY .

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
olvasás	Reál szakosok	10	3,47	,62	,20	3,02	3,91	2,49	4,67
	Humán szakosok	10	4,76	,94	,30	4,09	5,43	2,93	6,06
	Művészeti szakosok	10	4,66	1,09	,35	3,88	5,44	3,11	6,21
	Total	30	4,30	1,06	,19	3,90	4,69	2,49	6,21

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
olvasás	2,53	2	27	,098

I. lépés: ellenőrizzük, hogy p > .05. Ha ez nem teljesül, akkor nem-parametrikus alternatívát kell válasszunk!

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
olvasás	Between Groups	10,33	2	5,17	6,28	,006
	Within Groups	22,21	27	,82		
	Total	32,55	29			

Első lépésünk a második táblázat szemrevételezése: *Test of Homogeneity of Variances*. Ez a *Levene teszt* arra utal, hogy a csoportok közötti variancia azonos-e. Amennyiben a teszt eredménye szignifikáns, akkor a variancia nem azonos, és nem-parametikus alternatívát kell keresnünk az ANOVA helyett, mint pl. a *Kruskal-Wallis teszt*; amennyiben a teszt eredmény nem szignifikáns (p>.05), akkor az ANOVA teszt csoportok közötti egyenlő varianciára vonatkozó feltétele teljesült, és továbbléphetünk az eredmények értelmezésével.

Második lépésként a bekeretezett adatokat nézzük meg. Az első táblázatban *(Descriptives)* láthatjuk a csoportonkénti átlagot, míg az utolsó táblázatban az ANOVA teszt eredményét: az *F* és *p* értékeket. Mivel $p < .01$ (a teszt statisztikailag szignifikáns), elvetjük a nullhipotézist és megállapítjuk, hogy a csoportok átlagai különböznek. Most nézzük meg újra az átlagokat az első táblázatban: A reál szakosok $M = 3.47$-es átlag pontszáma a legkisebb, tehát ők töltenek a legkevesebb időt olvasással. Ha megnézzük a humán és művészeti szakosok átlagait, ($M = 4,76$ és $M= 4,66$), mindössze pár tized az eltérés. Ebből még nem tudjuk megállapítani, hogy a pár tizedes különbség statisztikailag szignifikáns-e (feltehetően nem lesz az), vagyis nem csak a véletlennek tudható be, ezért szükséges a harmadik lépés, a post-hoc teszt eredményeinek értelmezése is, amit a harmadikként megjelenített *(Multiple comparisons)* táblázatból tudhatunk meg:

Multiple Comparisons (olvasás)

	(I) Csoportok	(J) Csoportok	Mean Difference (I - J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	Reál szakosok	Humán szakosok	-1,29	,41	,010	-2,30	-,29
		Művészeti szakosok	-1,19	,41	,018	-2,20	-,19
	Humán szakosok	Reál szakosok	1,29	,41	,010	,29	2,30
		Művészeti szakosok	,10	,41	,967	-,91	1,11
	Művészeti szakosok	Reál szakosok	1,19	,41	,018	,19	2,20
		Humán szakosok	-,10	,41	,967	-1,11	,91

Mint láthatjuk, a *Tukey* próba páronként vizsgálja a csoportjaink közötti különbségeket.

Most tekintsük meg a szignifikancia szinteket: ahol $p < .05$, szignifikáns különbségek állnak fenn a csoportok között, ahol $p > .05$, a csoportok között nincs szignifikáns különbség.

A táblzat szerint:

– a reál és a humán szakosok csoportját párba állítva, a teszt eredménye szignifikáns ($p = .01$), tehát ezen csoportok átlaga különbözik;

– a reál és a művészeti szakosok csoportját párba állítva, a teszt eredménye szignifikáns ($p = .018$), tehát ezen csoportok átlaga is különbözik;

– humán és a művészeti szakosok csoportját párba állítva, a teszt eredménye nem szignifikáns

($p = .967$), ami arra utal, hogy a két csoport átlaga nem különbözik.

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

Egytényezős csoportok[44] közötti ANOVA teszttel vizsgáltuk a reál, humán és művészeti szakos diákok pontszámait, az olvasásra szánt idő tekintetében. A *Levene féle teszt* eredménye nem szignifikáns ($p > .05$), így a variancia homogenitásának feltétele teljesült, ezért a teszt alkalmazható. A csoportok átlag pontszámai között különbség van: $F(2, 30) = 6{,}28$, $p < .01$. A csoportok páronkénti utólagos (post-hoc) vizsgálatát *Tukey HSD féle teszt*tel végeztük, mely szerint a reál szakosok átlaga ($M = 3{,}47$, $SD = .62$) szignifikánsan különbözik a humán ($M = 4.76$, $SD = .94$) és művészeti ($M = 4{,}66$, $SD = 1.09$) szakosok átlagától, valamint a humán és a művészeti szakosok átlaga között nem tapasztaltunk szignifikáns különbséget.

Jelölése:

M=4,66, SD =1.09

ahol *M* átlag, *SD* szórás

F(2, 30) = 6,28, p < .01

ahol *F* érték (szabadsági fokok (*df*), esetek száma (*n*)) = a teszt eredménye, *p* érték

44 Az “egytényezős csoport” egy olyan statisztikai kifejezés, amely egy adott változó összehasonlítását jelenti különböző csoportok vagy kategóriák között egyetlen faktor (vagy más néven tényező) alapján.

12.7. Változók közti kapcsolatok vizsgálata

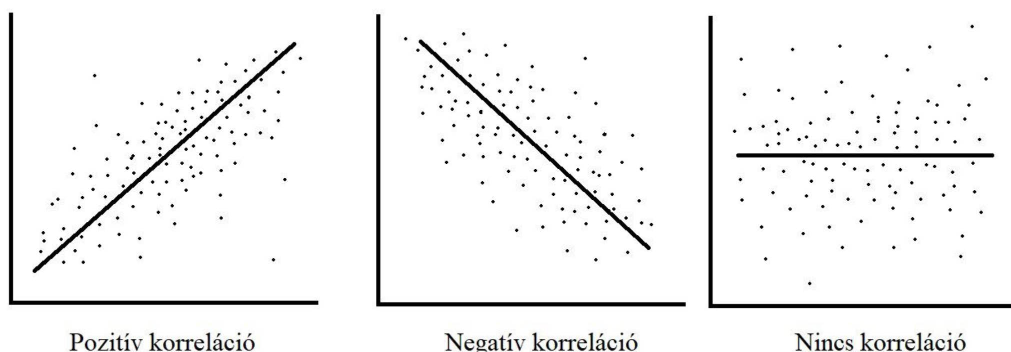
Az előző alfejezetekben (12.5. Nem parameters tesztek és 12.6. Paraméteres tesztek) a csoportok közötti különbségek viszgálatára alkalmas próbákat mutattam be; jelen alfejezetben a változók közötti kapcsolat vizsgálatára alkalmas, gyakran használt technikákról lessz szó. Ilyenek a korrelációs számítások, a főkomponens analízis és a regressziós analízis.

A korrelációk számítása viszonylag egyszerű, PSPP-ben is megvalósítható, a többi technika használatára – komplex jellegükből kifolyólag – az ingyenes statisztikai program már csak feltételesen ajánlott. A PSPP-ből hiányoznak bizonyos funkciók, vagy csak bonyolult parancssorokkal hajthatók vérge a *Syntax Editorban;* az eredmények értelmezése és bemutatása túl sok utómunkával jár. Ebben a tekintetben sz SPSS fölénye egyértelmű, ezért a korrelációs összefüggések vizsgálatán kívül a parancssorokat és a szükséges beállításokat már csak SPSS-ben fogjom ismertetni.

Általában az alap és mesterszakos dolgozatokhoz az egyetemek nem kérnek főkomponens analízist és a regressziós analízist. Ennek ellenére talán akad majd néhány hallgató, akiknek mégis hasznára lehet, ezért ezeket a technikákat is ismertetem.

12.7.1 Korrelációs összefüggések vizsgálata

A korrelációs összefüggések két változó együttmozgását mutatják ki, a köztük lévő lineáris kapcsolat nagyságát és irányát jelzik. Ez gyakorlatilag azt eredményezi, hogy az egyik változó értékeinek növekedése vagy csökkenése maga után vonja a másik változó értékeinek növekedését, illetve csökkenését. A korrelációs számításból nem tudunk arra következtetni, hogy mi okozza az összefüggést, csupán a kapcsolat meglétére utalhat. Az alábbi ábrán (9. sz.) szemléltetjük a korrelációs viszonyokat:



9. sz. Ábra – Korrelációs viszonyok

A korrelációs együttható jele „*r*”, lehetséges értékei (+1) és (-1) közé eshetnek, ahol (+1) a tökéletes együttmozgást, (0) a változók közti összefüggés hiányát és (-1) a tökéletes ellentétes mozgást jelöli. A korrelációs együttható értékei egyezményesen +/– 1-től +/– 0.5-ig erősnek, +/– 0.5-től +/– 0.3-ig közepesek, míg +/– 0.3-tól +/– 0.1-ig gyengének számítanak[45].

Például, pozitív korrelációs viszonyban lehet az életkor és a tanulmányi évek száma (vagyis az életkorral együtt növekszik az oktatás különböző formáiban eltöltött évek száma), negatív korrelációs viszonyban lehet felnőtteknél a növekvő testsúly és az állóképesség (vagyis a növekvő testsúly az állóképesség csökkenéséhez vezet). A jövedelem és a testsúly között nag valószínűséggel nem mutatkozik semmilyen összefüggés.

A teszt elvégzése során minden változó között tapasztalható egy bizonyos erősségű korrelációs együttható, viszont számunkra csak akkor értelmezhető ez az érték, ha a kapcsolat nem a véletlennek tudható be, vagyis statisztikailag szignifikáns (p<.01). Ezért a korrelációs együtthatók bemutatásakor mindig meg kell adnunk a szignifikancia szintet is.

Számítása viszonylag egyszerű, a korrelációs mátrix könnyen értelmezhető, viszont itt is figyelnünk kell a változóink típusára.

Két féle korrelációs számítási lehetőségünk van, az egyik a *Pearson féle korreláció* (r_P), a másik a *Spearman féle korreláció* (r_S). Normális eloszlású folytonos változók esetén Pearson féle korrelációt számolunk, nem-normális eloszlású változók és ordinális (rangsor) típusú változók esetén Spearman féle korrelációt. A kategorikus változók vizsgálatára a kereszttábla elemzést és a Khi-négyzet tesztet javasoljuk.

Parancssor és beállítások:
SPSS: Analyze → Correlate → Bivariate

A *Correlation Coefficients* beállításoknál kiválasztható, hogy Pearson vagy Spearman féle együtthatót szeretnénk számolni. A *Flag significant correlations* opciót is érdemes kiválasztani, így a korrelációs matrixban az *r* együttható után az SPSS (*) vagy (**) jelöléssel hívja fel a figyelmünket, ha $p < .01$ (vagyis a korrelációs viszony szignifikáns).

---

45 Egy másik megközelítés szerint a korrelációs együttható erősségének megítélése tovább bontható:
– 0: nincs lineáris kapcsolat;
– 0 - 0.2 (–0.2 - 0) : gyenge, majdnem hanyagolható kapcsolat;
– 0.3 - 0.4 (–0.4- –0.3) : biztos, de gyenge kapcsolat;
– 0.4 - 0.7 (–0.7 - –0.4) : közepes korreláció, jelentős kapcsolat;
– 0.7 - 0.9 (–0.7 - –0.9) : magas korreláció, markáns kapcsolat;
– 0.9 - 1 (–1 - –0.9) : nagyon magas korreláció, erős függő kapcsolat.

PSPP: Pearson és Spearman féle együttható is számolható, de más-más módon érhető el.

<u>– Pearson féle korreláció:</u>

Analyze → Bivariate Correlations

A *Flag significant correlations* opciót is érdemes kiválasztani a fent említett megfontolásból. Jelölése a PSPP-ben nem (*) vagy (**) lesz, hanem ($_{a}$).

<u>– Spearman féle korreláció:</u>

Közvetlen parancssort a PSPP nem tartalmaz, viszont a kereszttáblák beállításakor választhatjuk a *Spearman féle korrelációt*.

Analyze → Descriptive Statistics → Crosstabs

A számítás kissé körülményesebb, mint az SPSS-ben, mivel a változókat páronként adhatjuk meg (egy változót vagy a táblázat soraiba, vagy az oszlopba tehetjük be). A *Statistics* fül alatt kiválasztjuk a *Corr* opciót. A kimeneti ablak utolsó táblázatából leolvashatjuk a Spearman féle együtthatót:

Symmetric Measures

		Value	Asymp. Std. Error	Approx. T
Ordinal by Ordinal	Spearman Correlation	,92	,03	8,32
Interval by Interval	Pearson's R	,89	,04	6,92
N of Valid Cases		14		

Az eredmény értelmezése:

Ha a teszt eredménye nem szignifikáns ($p > .05$), akkor fenntartjuk a nullhipotézist és megállapítjuk, hogy a csoportok átlagai nem különböznek.

Ha a teszt eredménye szignifikáns ($p < .05$), akkor elvetjük a nullhipotézist és megállapítjuk, hogy a csoportok átlagai különböznek.

Magyarázat egy példán keresztül:

Példaként a Covid-19 alatt készült egyik felmérésünk adatbázisát használjuk. A kutatásban Likert skálán mértünk több olyan változót, amelyek az alanyok fizikai állapotára, családdal és barátokkal való viszonyokra, érzelmi jólétre, funkcionális jólétre és érzelmi életre vonatkoztak, a kijárási tilalom alatt. A korreláció számítás során megtudhatjuk, hogy ezen változók között van-e szignifikáns együttmozgás.

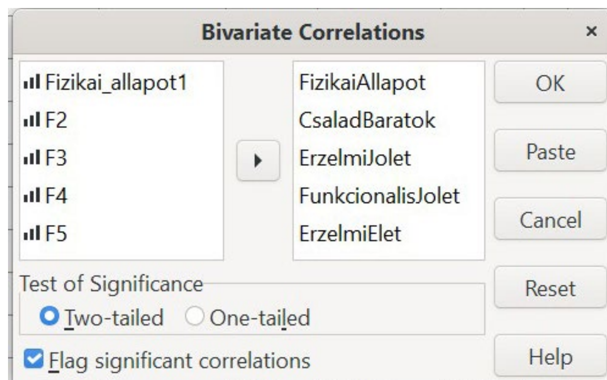
Az első lépésünk, hogy eldöntjuk, Pearson vagy Spearman féle együtthatót kell-e számolnunk. A normális eloszlású adatok esetén a Pearson féle együtthatót számoljuk, míg a nem-normális adateloszlás esetén a Spearman féle együtthatót. Ha még nem ismerjük az adataink eloszlását, akkor lefuttatunk egy Kolmogorov-Smirnov (KS) vagy egy Saphiro Wilk tesztet (a részleteket lásd a normalitásvizsgálatra vonatkozó 11.2.9. sz. alfejezetben).

One-Sample Kolmogorov-Smirnov Test

		Fizikai állapot	Család és barátok	Érzelmi Jólét	Funkcionális Jólét	Érzelmi Élet
N		14	14	14	14	14
Normal Parameters	Mean	1,36	2,24	1,83	2,60	5,29
	Std. Deviation	,79	,42	,50	,71	1,01
Most Extreme Differences	Absolute	,19	,14	,17	,25	,17
	Positive	,19	,14	,17	,18	,10
	Negative	-,12	-,14	-,16	-,25	-,17
Kolmogorov-Smirnov Z		,71	,54	,62	,95	,64
Asymp. Sig. (2-tailed)		,696	,931	,837	,331	,806

A KS teszt szerint az öt változónk eredménye közül egyik sem szignifikáns, ami normális eloszlásra utal (a KS teszt nullhipotézise az, hogy az adatok eloszlása szignifikánsan különbözik a normális eloszlástól; mivel a teszt eredménye nem szignifikáns, ezt a hipotézist elvetjük). Ebből az következik, hogy Pearson féle korrelációs együtthatót számolunk.

Az összes vizsgálni kívánt változónkat betesszük a *Bivariate Correlations* ablak üres mezejébe. Bejelöljük a *Flag significant correlations* opciót (lásd az indokot fent). Ezután az OK gombbal lefuttatjuk a tesztet.



A kimeneti ablakban a következő táblázatot kaptuk, amelyet *korrelációs mátrix*nak nevezünk:

Correlations

		Fizikai állapot	Család és barátok	Érzelmi Jólét	Funkcionális Jólét	Érzelmi Élet
Fizikai állapot	Pearson Correlation	1,000	,496	,894a	-,576a	-,530
	Sig. (2-tailed)		,071	,000	,031	,051
	N	14	14	14	14	14
Család és barátok	Pearson Correlation	,496	1,000	,602a	-,254	-,252
	Sig. (2-tailed)	,071		,023	,381	,386
	N	14	14	14	14	14
Érzelmi Jólét	Pearson Correlation	,894a	,602a	1,000	-,429	-,566a
	Sig. (2-tailed)	,000	,023		,126	,035
	N	14	14	14	14	14
Funkcionális Jólét	Pearson Correlation	-,576a	-,254	-,429	1,000	,528
	Sig. (2-tailed)	,031	,381	,126		,052
	N	14	14	14	14	14
Érzelmi Élet	Pearson Correlation	-,530	-,252	-,566a	,528	1,000
	Sig. (2-tailed)	,051	,386	,035	,052	
	N	14	14	14	14	14

a. Significant at .05 level

Az átló nem része a PSPP táblázatnak, viszont didaktikai célzatból berajzoltuk, mert segít eligazodni: tetszőlegesen vagy az átló alatti, vagy az átló feletti értékeket kell megvizsgáljuk (mivel az átló két oldala egymásnak tükörépe). Sőt, amikor az táblázatot áttesszük a dolgozatba, azokat a cellákat, amiket érint az átló, illetve a fölötte lévőket is kitöröljük, így egy kisebb méretű és áttekinthetőbb mátrixot kapunk.

A táblázatot úgy olvassuk, hogy a sorok és oszlopok metszéspontjában lévő cellák az adott sorban és az adott oszlopban lévő változók korrelációs együtthatóját (*r*) – első sor – , szignifikancia szintjét (*p*) – második sor –, és eseti számát (*n*) – harmadik sor – adják meg. Továbbá a korrelációs együtthatók után az ($_a$) jelölés szignifikáns összefüggésre utal.

Az eredmények megjelenítése a dolgozatban (a példa alapján**):**

A korrelációs mátrixot általában be kell másolni (a fent említett módon, az átló fölötti tükörképben szereplő cellákat törölve), majd szöveges formában minden szignifikáns összefüggést egyenként leírni:

A fizikai állapot szignifikáns és erős pozitív korrelációs viszonyban van az érzelmi állapottal ($r_p = .894$, $p < .05$, $n = 14$).

vagy

A funkcionális jólét és a fizikai állapot között fordított szignifikáns összefüggés figyelhető meg, a korrelációs együttható negatív előjelű és közepes erősségű ($r_p = -.576$, $p < .05$, $n = 14$).

Jelölése:

r = .03, p < .05; n=14

ahol r korrelációs együttható = értéke [-1, +1] tartományban lehet;
p szignifikancia szint; n az esetek száma

12.7.2. Főkomponens analízis

A főkomponens analízis (*Principal Component Analysis*, röviden *PCA*) egy adattömörítési technika, amely során több változót próbálunk kevesebb dimenzióba (kisebb számú változóba) süríteni, a köztük lévő korrelációk alapján. Olyankor alkalmazhatjuk, amikor nagyszámú, nehezen áttekinthető változónk van, és a bennük foglalt értékeket szeretnénk kevesebb számú, összevont változókban (faktorokban vagy főkomponensekben) megragadni, a köztük lévő kapcsolat erőssége (kötősédek) alapján. Az adattömörítés során új változókat hozunk létre, amelyeket faktornak, főkkomponensnek vagy háttér változónak is hívunk.[46]

A főkomponens analízist három lépésben végezhetjük el.

Első lépés: megvizsgáljuk, hogy a változóink alkalmasak-e az adattömörítésre; ez a minta méretétől és a változók közötti faktorsúly[47] (*loading*) erősségétől függ.

A minta minimum 150 esetet kell tartalmazzon, feltéve, ha erős faktorsúlyú változóink vannak, egyébként a 300-as esetszám javasolt, valamint szükséges az egyes változó párok között minimum r > .03 erősségű korrelációs összefüggés (Pallant, 2007).

A minta és a faktorsúlyok megfelelőségét két teszttel vizsgáljuk:

1. A *Bartlett-féle szférikussági teszt*, melynek szignifikánsnak kell lennie (p < .05).

2. A *Kaiser-Meyer-Olkin index* (röviden KMO), melynek .06 fölöttinek kell lennie egy 0 és 1 közötti tartományban.

Második lépés: a faktorkiválsztás, melynek célja a lehetséges legkisebb számú faktor meghatározása, a változók halmaza közötti összefüggések legjobb reprezentálása érdekében, ezáltal maximalizálva a faktorok varianciáját. A faktorkiválasztás során az alábbi szempontokra támaszkodhatunk:

A *Kaiser kritérium* alkalmazása során javasolt azon faktorok kiválasztása, melyeknek sajátértéke nagyobb, mint 1. A sajátérték (*eigenvalue*) az összes vál-

46 A főkomponenes analízis a faktoranalízis technikák egyike, tehát a főkomponens analízis fogalomköre szűkebb, egy konkrét tehnikára utal, míg a faktoranalízis fogalomköre általános, mely különböző technikákra utal. A faktor szót a statisztikában több értelemben is használjuk, a főkomponens elemzés során egy összevont változóra, míg a varianciaelemzés során a független változóra utal.

47 A faktorsúly a faktorok és az őket alkotó változók közötti korrelációt fejezik ki.

tozó varianciájából magyarázott variancia. Ha ez az érék kisebb, mint 1, akkor a faktorunk információtartalma kisebb, mint az eredeti változóké.

A *Cattell féle könyökszabály* (*Cattell's Scee Plot*) a sajátértékeket ábrázolja egy x és y tengelyen, majd a pontokat összeköti. Az összekötött pontok egy görbét rajzolnak ki. A görbe meredekségének hirtelen változása (töréspontja) fölötti sajátértékekkel rendelkező faktorokat érdemes kiválasztani. Elképzelhető, hogy ez a töréspont már az 1 alatti sajátértékekkel rendelkező faktoroknál megjelenik, így megfontolandó ezen faktoroknak is a kiválasztása, bár ilyenkor számolni kell az eredeti változókhoz képest lecsökkent információtartalmukkal.

A *Horn féle párhuzamos elemzés* során véletlenszerűen generálunk egy, a mintánkkal megegyező méretű adathalmazt. Összehasonlítjuk a mintánk faktorainak sajátértékét a véletlenszerű adatbázis faktorainak sajátértékével. Az elemzés következtében csak azokat a sajátértékeket tekintjük kiválasztásra érdemesnek, melyek meghaladják a véletlenszerű adathalmazban szereplő faktorok sajátértékét.

Jelen kézikönyvben csak a *Kaiser kritérium* és a *Catell féle könyökszabály* alkalmazására fogunk kitérni.

Harmadik lépés: a faktorok elforgatása, rotációja az értelmezésüket segítik. A rotáció lehet derékszögű (ortogonális) vagy ferde (oblikus), sőt ezen belül többféle eljárás is ismert. A rotáció típusának kiválasztásakor érdemes kipróbálni mindkét leggyakrabban használt eljárást: a Varimax derékszögű rotációt és a Direct Oblimin ferde rotációt. Az eredményeket a legkönnyebben értelmezhető struktúra szerint mutatjuk be.

A *Varimax eljárást* olyankor célszerű alkalmazni, amikor a faktorok között nincs korrelációs viszony, az *Oblimin rotációt* akkor javasolt alkalmazni, amikor a faktorok között bizonyos erősségű korrelációs összefüggések állnak fenn.

Ha a faktorok elemei közötti összefüggések mintázata egyértelmű, a különböző rotációs eljárások eredményei általában nagyon hasonlóak (Pallant, 2007). Tabachnick és Fiddell (2013, 651. o.) szerint "ha a korrelációk meghaladják az $r > .32$ értéket, akkor 10%-os (vagy több) átfedés van a faktorok közti varianciában, ami elegendő a ferde rotációhoz", például a *Direct Oblimin* alkalmazásához.

A rotációs eljárás kihatással van az egyes faktorok sajátértékeinek módosulására és az általuk faktoronként magyarázott varianciára, viszont az összes faktor által magyarázott teljes variancia aránya nem változik.

A társadalomtudományi kutatásokban áltlában a kiválasztott faktorok között van egy bizonyos erősségű korrelációs viszony, ezért az elemzést javasolt a *Direct Oblimin* rotációs eljárással kezdeni.

A főkomponens elemzés célja új főkomponensek (faktorok) létrehozása az adatbázisunkban szerepelő változók halmazából. Azt, hogy melyik változókat érdemes az analízisnek alávetni, a faktorsúlyuk alapján dönthetjük el. Az ala-

csony faktorsúlyú változók (<.03) kihagyása megfontolandó (a *Communalities* táblázatban ellenőrizhetjük).

Szintén eldöntendő, hogy szám szerint hány főkomponenst, (faktort) hozzunk létre a változó halmazunk alapján. Érdemes először megnézni, hogy 1-nél nagyobb sajátértékkel hány faktor jöhet létre (*Extraction* fül alatt *Based on Eigenvalue*). Mindegyik faktorról el kell dönteni, hogy a vele faktorsúlyuk szerint kötődő változók egy bizonyos logikai szempontból csoportosíthatóak-e, mivel később tartalmuk alapján nevezhetjük el őket. Ha szükséges, csökkenthetjük is a faktorok (főkomponensek) számát (*Extraction* fül alatt *Fixed number of factors/Factors to extract*), viszont ezáltal a faktorok által magyarázott teljes variancia is csökkeni fog.

Mivel egy adatfeltáró technikáról van szó, a faktorok kialakítása és az eredmények értelmezése szubjektív, főként a kutató megítélésén múlik.

Feltételei:

– egy egymással korreláló (leghetőleg $r > .03$ erősségű) változó halmaz (lineáris kapcsolat a változók között);

– a hamazt képező változók folytonosak;

– minimális mintaméret: 150 alany;

– a Bartlett féle szférikus teszt szignifikáns ($p < .05$) és a *KMO* értéke > .06.

Parancssor és beállítások:

Mivel a **PSPP** műveleti lehetőségei korlátozottak, csak az **SPSS**-re vonatkozóan adjuk meg a beállításokat.

SPSS: Analyze → Dimension Reduction → Factor

Az elemezni kívánt változóink halmazát betesszük a *Variables* mezőbe. A *Descriptives* fül alatt bejelöljük a következőket: *Initial solutions, Coefficients, KMO and Bartlett's test of sphericity*.

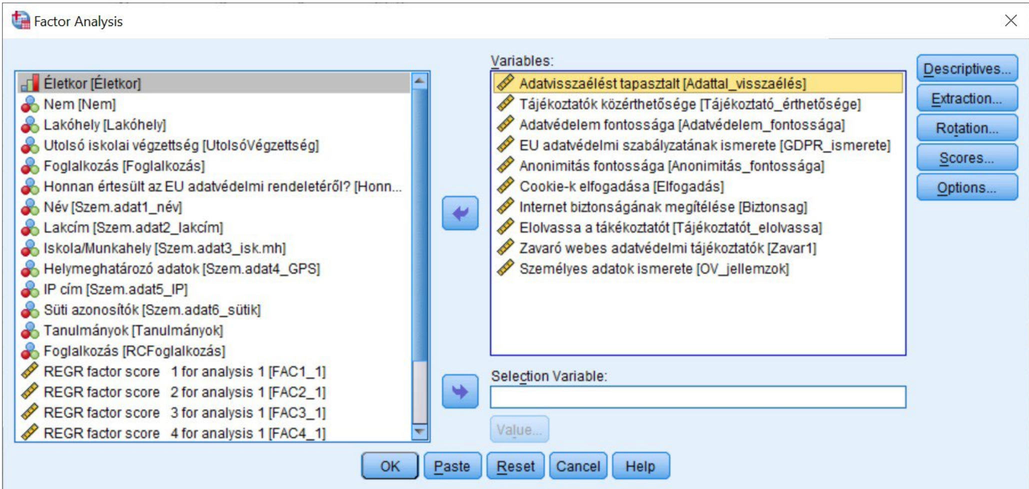
Az *Extraction* fül alatt *Principal components, Correlation matrix, Based on Eigenvalue, Unrotated factor solution, Scree plot; Maximum iterations for convergence:* 25. A *Rotation* fül alatt kiválasztjuk a *Direct Oblimin* metódust, *Delta = 0, Rotated Solutions, Maximum iterations for convergence: 25.* A *Scores* fül marad alapértelmezett. Az *Options* fül alatt: *Exclude cases pairwise, Supress small coefficients:* megadjuk .03-at vagy .04-et. Ez utóbbi funkció segíthet az analízis során a faktorok strukturális és mintázati táblázataiban való eligazodásban (kevesebb érték fog megjelenni), de az eredmények bemutatásakot ki kell kapcsolni, mivel az összes korrelációs és faktorsúly értéket meg fog kelleni adni egy táblázatban.

A műveletet többször is lefuttathatjuk különböző változók és faktorszámok kombinációival, egészen addig, amíg számunkra értelmezhető faktorokat nem kapunk.

Az eredmény értelmezése példán keresztül:

Példaként egy korábbi kutatásunk adatbázisát hoznám fel, amelyben 162 alany ismereteit és véleményét mértük fel az Európai Unió adatvédelmi rendeletével (GDPR) kapcsolatosan.

Összesen 10 *Likert skála* típusú változó képezi az elemzésre szánt változók halmazát. Ezeket betesszük a *Variables* mezőbe, majd elvégezzük a szükséges beállításokat a *Parancssor és beállítások* alcím szerint.



Első lépés: Ellenőriztük a KMO és Bartlett féle tesztek feltételeinek teljesülését, amit rendben találtunk, megállapítottuk, hogy az adatbázis változói alkalmasak az adattömörítésre (*KMO* = .66, Bartlett féle teszt szignifikáns $p < .01$). A korrelációs mátrixban (ez a legelső táblázat) több $r > .03$ erősségű összefüggést is találtunk.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,656
Bartlett's Test of Sphericity	Approx. Chi-Square	161,681
	Df	36
	Sig.	,000

Második lépés: Megnézzük, hogy hány faktoros megoldást kaptunk, illetve a faktorok között milyen korrelációs viszony áll fenn.

Component Correlation Matrix

Component	1	2	3
1	1,000	-,090	,157
2	-,090	1,000	-,135
3	,157	-,135	1,000
Extraction Method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalization.			

Jelen példában három főkomponenst azonosítottunk, a *Component Correlation Matrix* táblázat szerint a faktorok közötti korreláció nagyon alacsony ($r = .157$; -135; -.09). Ezek az értékek kisebbek a Tabachnick and Fiddell (2013) általl javasolt $r > .32$ szintnél, amely a ferde rotációhoz szükséges. Ezért megismételjük az analízist, de ezúttal a *Direct Oblimin* rotációs módszer helyett *Varimax* típusú rotációt alkalmazunk.

Az ismételten lefuttatott művelet során szintén 3 faktoros megoldást kapunk, amelyek a teljes variancia 54,59%-át magyarázza (ez nem függ a rotáció típusától). A három faktor sajátértékét és a magyarázott variancia értékeket a *Total Variance Explained* táblázatból (*Cumulative* %) olvashatjuk ki. A több faktoros megoldás a teljes variancia nagyobb százalékát, míg a kevesebb faktoros megoldás a teljes variancia kisebb százalékát magyarázná. A táblázatban megkaptuk a faktoronként magyarázott variancia százalékát is, a teljes varianciához viszonyítva (% *of Variance*).

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,172	24,138	24,138	2,172	24,138	24,138	1,752	19,465	19,465
2	1,553	17,257	41,395	1,553	17,257	41,395	1,746	19,398	38,863
3	1,188	13,201	54,595	1,188	13,201	54,595	1,416	15,733	54,595
4	,889	9,875	64,470						
5	,795	8,836	73,306						
6	,684	7,605	80,910						
7	,628	6,973	87,883						
8	,605	6,718	94,601						
9	,486	5,399	100,000						

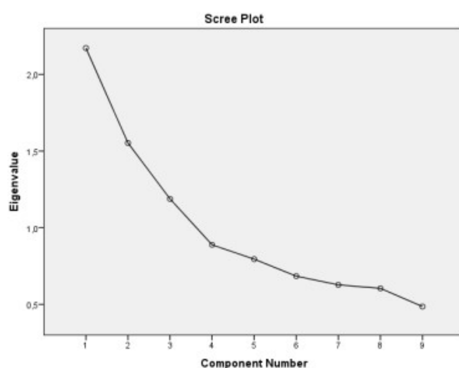
Extraction Method: Principal Component Analysis.

A *Communalities* táblázatban azt ellenőrizhetjük, hogy az egyes változóinkat milyen arányban magyarázzák a faktorok. Az alacsony kommunalitási értékek (<.03) arra utalnak, hogy az adott változó nem illeszkedik jól a többi változóhoz képest a faktorban. Ha a nem illeszkedő elemeket kivesszük az analízisből, növelhetjük a faktorok által magyarázott teljesvarianciát (Pallant, 2007). Az eredmények dolgozatban (tanulmányban) való megjelenítésekor ezeket az értékeket is meg fog kelleni adnunk.

Communalities

	Initial	Extraction
Adatvisszaélést tapasztalt	1,000	,538
Tájékoztatók közérthetősége	1,000	,364
Adatvédelem fontossága	1,000	,590
EU adatvédelmi szabályzatának ismerete	1,000	,628
Anonimitás fontossága	1,000	,643
Cookie-k elfogadása	1,000	,401
Internet biztonságának megítélése	1,000	,539
Elolvassa a tákékoztatót	1,000	,678
Személyes adatok ismerete	1,000	,534
Extraction Method: Principal Component Analysis.		

Most megtekintjük a könyökszabály (*Scree plot*) ábráját, és azt elemezve megállapítjuk, hogy a görbe meredeksége már a 4. főkomponenstől kezd hirtelen emelkedni, viszont csak háromnak nagyobb a sajátértéke, mint 1. Ez arra utal, hogy a három komponenses megoldás megfelelő lehet.



A Varimax rotációs módszerrel további két táblázatot kapunk: az egyik a *Component Matrix*, a rotáció előtti faktorsúlyok értékeit, a másik a *Rotated Component Matrix,* a rotáció utáni faktorsúly értékeket tartalmazza. A dolgozatban a *Rotated Component Matrix* adatait szükséges megadni (a *Component matrix* mehet a függelékekhez), ami a példánkban így néz ki:

Rotated Component Matrix[a]

	Component		
	1	2	3
Adatvisszaélést tapasztalt	**,653**	-,091	-,322
Tájékoztatók közérthetősége	-,318	**,483**	,170
Adatvédelem fontossága	,125	,129	**,747**
EU adatvédelmi szabályzatának ismerete	,329	**,696**	-,189
Anonimitás fontossága	**,719**	,052	,351
Cookie-k elfogadása	-,136	**,618**	,028
Internet biztonságának megítélése	-,218	,022	**,700**
Elolvassa a tákékoztatót	-,109	**,772**	,264
Személyes adatok ismerete	**,711**	-,150	-,077

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 4 iterations.

A táblázatban kiemeltük a .5 körüli és az ennél nagyobb faktorsúlyokat, mivel a faktor (főkomponens) ezeket ragadja meg leginkább a faktorsúlyuk alapján, másképpen fogalmazva ezek a változók jellemzik a leginkább a faktort.

A három faktor a fenti táblázat alapján a következő:
1. Faktor: adatvisszaélés tapasztalása, anonimitás fontossága, személyes adatok ismerete.
2. Faktor: adatvédelmi tájékoztatók közérthetősége, EU adatvédelmi szabályzatának ismerete, Cookie-k tudatos elfogadása, tájékoztatók elolvasása.
3. Faktor: adatvédelem fontossága, internet biztonsága.

Most megpróbálhatjuk logikai szempont szerint értelmezni és elnevezni a főkomponenseket. Abból indulunk ki, hogy milyen értelem szerinti kapcsolat lehet ezen változók között:
1. Faktor: a mi olvasatunkban egyfajta érzékenységet jelez, tehát nevezhetjük *adatvédelmi érzékenységi mutatónak*.
2. Faktor: véleményünk szerint egyfajta *adatvédelmi tudatosságra* utal.
3. Faktor: úgy gondoljuk, hogy ez a faktor *az adatvédelem fontosságának megítélésére* utalhat.

Tehát az eredeti 10 változóból az adatvédelmi ismeretek és attidűdök három háttérváltozón keresztül is megragadhatók. A három faktor a teljes variancia 54,59%-át magyarázza, ebből az 1. számú faktor az *adatvédelmi érzékenységet* tartalmazza (a teljes variancia 24,1%-át magyarázza), a 2. számú faktor az *adatvédelmi tudatosságot* tükrözi (a teljes variancia 17,2%-át magyarázza), a 3. sázmú faktor az *adatvédelem fontosságára* irányul (a teljes variancia 13,2%-át magyarázza).

Amennyiben *Direct Oblimin* rotációs eljárást alkalmazzuk (ha a komponensek egymás között erősebben korreláltak volna), a művelet során három táblázatot kapunk:

Component Matrix: a rotáció előtti faktorsúlyokat tartalmazza.

Pattern Matrix: az *Oblimin* rotáció utáni faktorsúlyokat tartalmazza.

Structure Matrix: az eredeti változók és a faktorok közötti korrelációkat tartalmazza.

Az eredmények bemutatásakor a *Component Matrix*-ot a dolgozat Függelékek részében szokás feltüntetni, a *Pattern*- és *Structure Matrixokat* a *Communalities* értékekkel együtt egy összevont táblázatban szükséges feltüntetni a szövegtestben.

Az eredmények megjelenítése a dolgozatban (a példa alapján):

A kutatásból kiemeltünk tíz változót, melyeken keresztül az adatvédelemmel kapcsolatos ismereteket és attitűdöket mértük, majd főkomponens analízissel megvizsgáltuk őket. Célunk – adattömörtési szempontból – olyan összetett változók kialakítása volt, melyekkel magyarázható a változók közötti variancia bizonyos hányada.

Az adatok műveleti alkalmasságát tekintve a változók közötti korrelációs mátrixban több olyan összefüggést találtunk, melyek erőssége .3 feletti (a szövegtestbe átmásoljuk a táblázatot). A minta nagysága megfelelő (n=162), a KMO értéke .6, a Bartlett-féle szférikus próba statisztikailag szignifikáns (p=.01), tehát az adataink alkalmasak a vizsgálathoz.

A főkomponens analízis során három összevont változót sikerült kialakítani a *Kaiser féle sajátérték kritérium*ra (a faktorok sajátértéke >1) és a *Cattell féle könyökszabály*ra támaszkodva (a görbe meredekségének hirtelen emelkedése alapján) (függelékben mellékeljük a *Component Matrix* táblázatot és a *Scree-plot* ábrát).

A három főkomponens a teljes variancia 54,59%-át magyarázzák (mellékeljük a *Total Variance Explained* táblázatot).

A faktorok értelmezését derékszögű Varimax rotáció segítette. Ezt a típusú faktorrotációs eljárást a főkomponensek közötti nagyon alacsony korreláció ($r \leq .157$)[48] indokolta (betesszük a szövegtestbe vagy a függelékekhez a *Component Correlation Matrix* táblázatot).

Az elemzés során a következő faktorokat azonosítottuk, melyek a kutatásunkban segítettek az adatvédelemmel kapcsolatos ismeretek és attitűdök faktorsúly alapú rendszerezéséhez:

48 Természetesen kerekítve jelenítjük meg: $r \leq .16$, csak a követhetőség érdekében didaktikai célzattal írtuk ki a pontos korrelációs együttható értékét.

Oblimin rotció esetén betesszük a *Pattern-* és *Structure Matrixokat, valamint* a *Communalities* értékeit.

Varimax rotáció esetén betesszük a *Rotated Component Matrix* és *Communalities táblázatokat.*

Az 1. számú faktor az *adatvédelmi érzékenységet* tartalmazza (a teljes variancia 24,1%-át magyarázza), a 2. számú faktor az *adatvédelmi tudatosságot* tükrözi (a teljes variancia 17,2%-át magyarázza), a 3. számú faktor az *adatvédelem fontosságára* irányul (a teljes variancia 13,2%-át magyarázza). Komponensekre bontva: 24,1%, 17,2% és 13,2%, valamint 11,7%. A főkomponenseket akár táblázatokba is rendezhetjük - lásd X. sz. Táblázat.

X. sz. Táblázat – A főkomponens analízis faktorai

Főkomponensek	**Erős kötődést mutatató komponensek**
1.	adatvisszaélés tapasztalása, anonimitás fontossága, személyes adatok ismerete
2.	adatvédelmi tájékoztatók közérthetősége, EU adatvédelmi szabályzatának ismerete, Cookie-k tudatos elfogadása, tájékoztatók elolvasása
3.	adatvédelem fontossága, internet biztonsága

Emlékeztető:

1. mindig figyeljünk arra, hogy fejezetet, alfejezetet ne táblázattal vagy ábrával zárjunk.

2. az APA hivatkozási stílusban a táblázatokban nincsenek függőleges vonalak, tehát a statisztikai program által generált táblázatokat ne feledkezzünk el formailag megfelelően átszerkeszteni.

12.7.3. Többszörös lineáris regresszió

A többszörös lineáris regressziós eljárás célja egy folytonos függő változó és több független változó közötti kapcsolat feltárása. öbbféle változata ismert, például a standard, hierarchikus és lépcsős regresszió. Az eljárás során megvizsgálhatjuk, hogy a független változók közül melyik a legjobb prediktor. A prediktorokat nevezzük független változóknak.

A módszer változók közötti lineáris függvénykapcsolatokat határoz meg egy deteminációs együttható alapján (r^2)[49]. A regresszió során egy olyan modellt hozunk létre, amely a változók összessége és az egyes változók hatásáról is képes előrejelző, un. prediktív információkat adni.

49 A determinációs együttható a korrelációs együttható négyzetre emelve.

Az is vizsgálható, hogy a hozzáadott változók befolyásolják-e a modell előrejelző képességét. A regresszió számítást befolyásolják a kiugró értékek (*outliers*), ezért ezeket szükséges lehet törölni vagy átalakítani.

Ebben a kézikönyvben a leggyakrabban használt standard többszörös regresszióra térünk ki.

A standard többszörös regresszió számításánál egyszerre adjuk meg a függő és független változókat, majd vizsgáljuk a független változók prediktív erejét, és meghatározhatjuk, hogy mely független változók magyarázzák az egyedi varianciát a függő változóban.

A standard többszörös regresszió elsősorban olyan kutatásokban alkalmazható, amelyekben nagyszámú és ténylegesen reprezentatív mintával dolgozunk, akár saját adatgyűjtésből, de főként másodlagos adatelemzési forrásokból.

Feltételei:

– egy függő változó, mely folytonos és normális eloszlású;

– kettő vagy több független változó (prediktor), amelyek lehetnek folytonosak és/vagy dihotóm változók;

– a minta nagysága: a szakirodalomban a mintaméretre vonatkozóan többféle javaslat is felemrült, elfogadható arány lehet a 15 alany/prediktor vagy 5 prediktorhoz minimum 90 eset (Pallant, 2007);

– előzetesen ellenőrizni kell a normalitás, linearitás és homoszkedaszticitás[50] feltételeinek teljesülését a reziduumokra[51] vonatkozóan.

Parancssor és beállítások:

Mivel a **PSPP** műveleti lehetőségei korlátozottak, alább csak az **SPSS**-re vonatkozóan adjuk meg a beállításokat.

SPSS: **Analyze → Regression → Linear**

A folytonos függő változónkat betesszük a *Dependent* mezőbe. A független változóinkat betesszük az *Independet* mezőbe. Az *Independent* mező alatt a *Method* marad az alapértelmezett *Enter* (ezt a metódust kell megadni a standard regresszióhoz). A *Statistics* fül alatt ki kell választani: *Estimates*, *Confidence Intervals*, *Model fit*, *Descriptives*, *Part and partial correlations* és *Collinearity diagnostics*. Ugyanitt a *Rediduals* beállításoknál meg kell adni: *Casewise diagnostics* és *Outliers outside* 3 (alapértelmezett) *standard deviations*. Az Options fül alatt *Missing Values: Exclude cases pairwise.* A *Plots* fül alatt kiválasztjuk a **ZRESID*-et az *Y* mezőbe, *a *ZPRED*-et az *X* mezőbe[52], majd bejelöljük a *Normal probability plot*-ot**.**

50 A reziduumok állandó varianciával rendelkeznek.

51 A reziduumok (hibatagok) a kapott és a megjósolt függő változó pontszámai közötti különbségek.

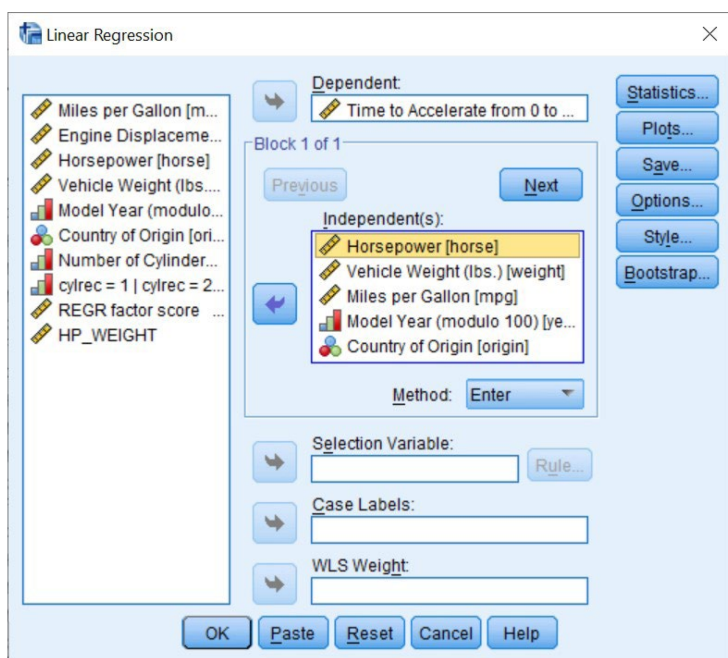
52 ZPRED a standardizált becsült értékre, ZPRESID a standardizált reziduumokra vonatkozik.

Az eredmény értelmezése és magyarázatok egy példán keresztül:

A példánk alapjául egy online elérhető adatbázist használunk a Michigani Egyetem online SPSS Training Workshop oldaláról.[53] A példánk a változatosság kedvéért műszaki jellegű, mivel nevezett adatbázis műszaki paramétereket listáz, de lehetne szó a társadalomtudományokban szokványos véleményekről, attitűdökről, magatartásokról is, tehát bármi, ami mérhető és amiről elegendő adattal rendelkezünk, képezheti vizsgálatunk tárgyát. Ebben az adatbázisban 406 autó szerepel bizonyos technikai adatokkal (ez a minta), mint például gyorsulás, tömeg, teljesítmény, fogyasztás, stb. (ezek lesznek a változók). A kutatás célja az autók gyorsulását befolyásoló tényezők feltárása, illetve a legerősebb prediktor meghatározása. A gyorsulás lesz a függő változónk, a teljesítmény, tömeg, fogasztás, gyártási év és gyártási hely lesznek a független változók. Amenyiben a kutatásban például az autók fogyasztását vizsgálnánk, akkor az lenne a függő változónk. Tehát a kutató mérlegelésén múlik a modell kialakítása.

A regresszió elemzés eredményének kiértékelése több lépésben történik; az adataink műveleti alkalmasságának ellenőrzésével kezdünk. Ezek a multikollinearitás, a normalitás, linearitás és a homoszkedaszticitás.

Az alábbiakban bemutatott példában látni fogjuk, hogy néhány kezdeti lépést meg kellett ismételnünk ahhoz, hogy a műveleti alkalmasság feltételei teljesüljenek, illetve a tervezett változók számát is módosítani kellett. A változókat megadjuk az alábbiak szerint, elvégezzük a szükséges beállításokat és lefuttatjuk az analízist.



53 Az adatbázis elérhető 2023.02.02-án: http://calcnet.mth.cmich.edu/org/spss/prj_carsdata.htm

Az első lépés *a multikollinearitás* vizsgálata, vagyis magas interkorrelációkat keresünk két vagy több független változó között. Segítségünkre van az eredmények kimeneti ablakában az első korrelációs táblázat (*Correlations*): akkor tekinthetjük ideálisan teljesültnek ezt a feltételt, ha a független változók között .03 és 0.7 közötti erősségű korrelációs összefüggéseket tapasztalunk. Amennyiben túl erős a korreláció két változó között (r > .07), csak az egyiküket érdemes bennehagyni a modellben, esetleg összevont faktort kialakítani belőlük.

Correlations

		Accel. from 0 to 60 mph (sec)	Horsepow.	Vehicle Weight	Miles per Gallon	Model Year	Country of Origin
Pearson Correlation	Accelerate from 0 to 60 mph (sec)	1,000	-,661	-,334	,372	,249	,176
	Horsepower	-,661	1,000	,864	-,780	-,349	-,458
	Vehicle Weight (lbs.)	-,334	,864	1,000	-,789	-,248	-,576
	Miles per Gallon	,372	-,780	-,789	1,000	,543	,546
	Model Year	,249	-,349	-,248	,543	1,000	,156
	Country of Origin	,176	-,458	-,576	,546	,156	1,000
Sig. (1-tailed)	Accelerate from 0 to 60 mph (sec)	.	,000	,000	,000	,000	,000
	Horsepower	,000	.	,000	,000	,000	,000
	Vehicle Weight (lbs.)	,000	,000	.	,000	,000	,000
	Miles per Gallon	,000	,000	,000	.	,000	,000
	Model Year (modulo 100)	,000	,000	,000	,000	.	,001
	Country of Origin	,000	,000	,000	,000	,001	.
N	Accelerate from 0 to 60 mph (sec)	383	377	383	376	382	382
	Horsepower	377	377	377	370	376	376
	Vehicle Weight (lbs.)	383	377	383	376	382	382
	Miles per Gallon	376	370	376	376	375	375
	Model Year (modulo 100)	382	376	382	375	382	382
	Country of Origin	382	376	382	375	382	382

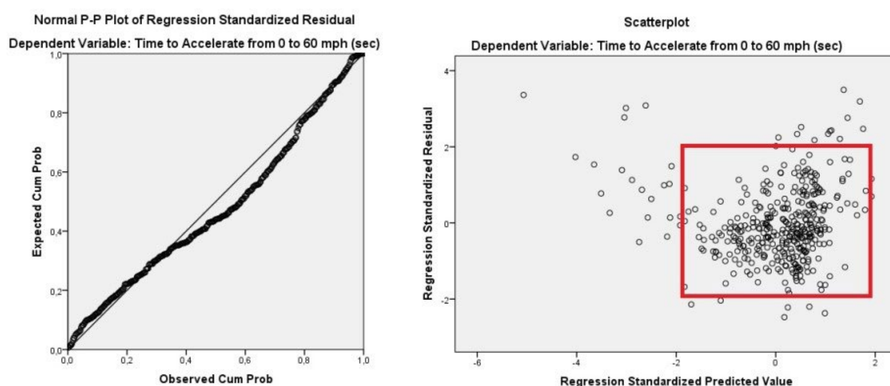
Az adatok forrása: Central Michigan University (s.a.).

A fenti táblázatban[54] megfigyelhetjük, hogy a teljesítmény és a tömeg között nagyon erős a korrelációs összefüggés (r = .864), valamint a teljesítmény és a fogyasztás között is (r = .78), ami a modell szempontjából a nem kívánatos tartományba esik (r>.07), ezért vagy eltávolítjuk a két változót, vagy összevonjuk őket. Ebben a példában – az egyszerűség kedvéért – eltávolítjuk a két változót, azzal a megjegyzéssel, hogy hasonló helyzetben, egy valós kutatás során, megoldást kellett volna találnunk a változók összevonására. Egyik lehetőség – ha két vagy több változónk van –, a *Transform* menüpontból a *Compute Variable* opció,

54 A fenti táblázatot már a kiugró értékek eltávolítása után kaptuk.

itt akár tetszőlegesen súlyozhatjuk is a változóinkat, feltéve, ha megtudjuk indokolni. Másik lehetőség – ha három vagy több változónk van (halmaz) –, a főkomponens analízis lett volna, amely során például a teljesítményből, tömegből és fogyasztásból létrehozhattunk volna egy faktort.

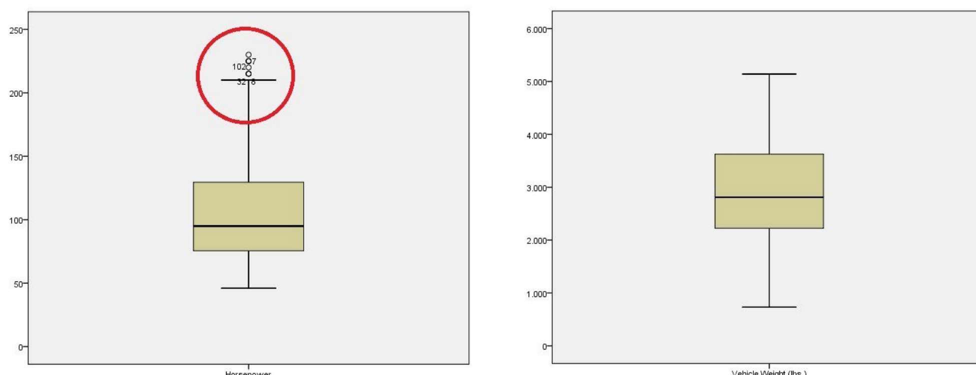
Második lépés a *normalitás, linearitás és homoszkedaszticitás* vizsgálata. Ezt a művelet során generált két diagramon is ellenőrizhetjük: az egyik a normál valószínűségi diagram (*Normal Probability Plot (P-P)*) a másik a szórásdiagram (*Scatterplot*). Az adataink megfelelőek lehetnek, ha a normál valószínűségi diagramon egy viszonylag egyenes, átlósan balról jobbra emelkedő vonalat látunk. A szórásdiagramon a maradványértékek nagyjából egy téglalap formájú alakzatba kellene tömörüljenek, amelyben a pontok többsége az alakzat közepén helyezkedik el (Pallant, 2007). A szórásdiagramon azonosíthatóak a kiugró értékek is, amelyeket onnan ismerhetünk fel, hogy szórási maradványértéke nagyobb 3,3-nál vagy kisebb –3,3-nál (Tabachnick és Fiddell, 2013, 128. o.).



A fenti ábrákon az látható, hogy a normál valószínűségi diagramon az átló mentén elhelyezkezkedő értékek nagyjából követik az átló vonalát, viszont a szórásdiagramon egyértelmű a kiugró értékek jelenléte. A keretes részben nincs egyértelmű mintázat, sőt a pontoknak megfelelő értékek is nagyjából a téglalap közepébe tömörülnek, ami az adatelemzés szempontjából megfelelő. Viszont, ha pontos eredményeket szeretnénk kapni, a kiugró értékeket törölnünk kell, kivéve, ha számuk minimális. A kiugró értékeket a független változók körében ellenőrizük, a következő módon:

SPSS: Analyze → Descriptive Statistics → Explore

A *Dependent List* mezőbe betesszük a vizsgálni kívánt változóinkat, a *Factor List*-et üresen hagyjuk, majd lefuttajuk a műveletet. Látszólag megtévesztő lehet, hogy ennél a műveletnél a regressziós elemzés szempontjából függetlennek tekintett változókat kell betenni a függő változók mezőbe, viszont most a célunk a kiugó értékek beazonosítása. Az alábbi dobozdiagramot elemzve azt látjuk, hogy a teljesítmény (*Horsepower*) változónk esetében vannak kiugró értékek, a többi független változónkál nem tapasztaltunk kiugró értékeket.



Az első dobozdiagram bekeretezett részében olvashatjuk le azon sorok számát, amelyek kiugró értékeket tartalmaznak, konrétan ezek a 7, 8, 32 és 102-es sorok lesznek az adatbázisunkban (lásd a piros körben kiemelt értékeket). A második, példaként felhozott dobozdiagramban nincsenek kiugró értékek.

A kiugró értékeket tartalmazó sorokat töröljük, majd újra lefuttatjuk az elemzést. Arra figyeljünk a törléskor, hogy először jelöljük ki az összes eltávolításra szánt sort (*Ctrl* billentyűt lenyomva tartva kijelöljük, majd az utolsó sárgával kiemelt soron job klikk és *Clear*), mert ha egyenként törölnénk, elcsúszna a sorok számozása. Az *Explore* műveletet ismételjük meg és ellenőrizzük a dobozdiagramot, hogy ne maradjanak kiugró értékek. Elképzelhető, hogy ez utóbbi műveletet többször is meg fog kelleni ismételnünk[55].

Miután eltávolítottuk a kiugró értékeket és, feltéve, ha a korrelációs táblázat együtthatói is a kívánt tartományban vannak, újra lefuttatjuk a regressziós elemzést, a fent már ismertetett beállításokkal. Ebben a konkrét példában a regressziós modelből kihagyjuk a tömeget, mert nagyon erős korrelációs összefüggést mutatott a teljesítménnyel. Most szükségessé válik az első lépés megismétlése, a korrelációs táblázat vizsgálata, mivel eltávolítottunk két változót és öt helyett már csak három független változóval számolunk. Egy valós kutatás során a magasan korreláló változók főkomponens alanízise lett volna a helyes eljárás, viszont mi didaktikai célzatból az egyszerűbb utat választottuk. Az új táblázat így néz ki:

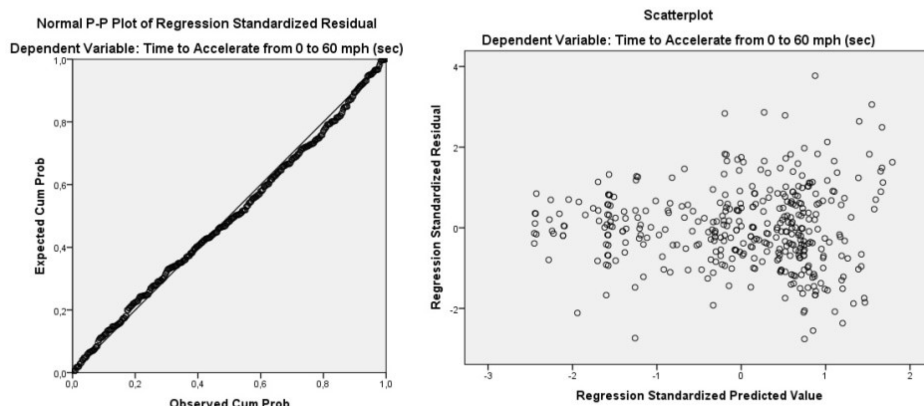
55 A kiugró értékek kezelésének több más módja is van, például a *Mahalanobis féle távolságok* alapján (lásd Pallant, 2007, 159. o), viszont jelen kézikönyvben ezt nem részletezük, mert a fenti módszert egyszerűbbnek találtuk.

Correlations

		Accelerate from 0 to 60 mph	Horsepower	Model Year	Country of Origin
Pearson Correlation	Time to Accelerate from 0 to 60 mph (sec)	1,000	-,661	,249	,176
	Horsepower	-,661	1,000	-,349	-,458
	Model Year (modulo 100)	,249	-,349	1,000	,156
	Country of Origin	,176	-,458	,156	1,000
Sig. (1-tailed)	Time to Accelerate from 0 to 60 mph (sec)	.	,000	,000	,000
	Horsepower	,000	.	,000	,000
	Model Year (modulo 100)	,000	,000	.	,001
	Country of Origin	,000	,000	,001	.
N	Time to Accelerate from 0 to 60 mph (sec)	383	377	382	382
	Horsepower	377	377	376	376
	Model Year (modulo 100)	382	376	382	382
	Country of Origin	382	376	382	382

A fenti táblázatban azt látjuk, hogy a korrelációs értékek nagyjából a javasolt tartományon belül vannak (r = .3 - .7).

Megismételjük a második lépést és ellenőrizzük a normál valószínűségi diagramot és a szórásdiagramot:



A normál valószínűségi diagram az elvártak szerint alakult, vagyis a normalitás feltétele teljesült. A szórásdiagram alakzata nagyjából elfogadható, bár nem tökéletesen szimmetrikus, a pontok centruma enyhén balra tolódott. Az alakzatról elmondható, hogy mindenképpen javult a kiugró értékek törlése után. Ez a regresszióban a maradvány értékek linearitásának hibájára utal, viszont nem érvényteleníti az elemzést, hanem gyengíti azt (Tabachnick és Fiddell, 2013, 127. o.).

Harmadik lépés a *regressziós modell értelmezés*e a *Model Summary* táblázat alapján. Az *R Square* oszlop megadja, hogy a függő változó hány százalék varianciáját magyarázza a modell (Az *R Square* értéket beszorozzuk 100-al). Az *Adjusted R square* érétke pontosabb, a valóshoz közelebb álló megközelítést ad kisebb mintás modellek esetén. Ekkor még nem tudjuk, hogy a modellünk statisztikailag szignifikáns-e, ennek megállapításához a következő táblázat (*ANOVA*) utolsó oszlopából kell kiolvasnunk a *p* értéket (szignifikáns, ha $p < .05$).

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,676[a]	,457	,453	1,986

a. Predictors: (Constant), Country of Origin, Model Year (modulo 100), Horsepower
b. Dependent Variable: Time to Accelerate from 0 to 60 mph (sec)

A példában szereplő regressziós modell a függő változó (gyorsulás) varianciájának 45,7%-át magyarázza.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1234,949	3	411,650	104,414	,000[b]
	Residual	1466,601	372	3,942		
	Total	2701,550	375			

a. Dependent Variable: Time to Accelerate from 0 to 60 mph (sec)
b. Predictors: (Constant), Country of Origin, Model Year (modulo 100), Horsepower

Az ANOVA teszt szignifikáns ($p<.01$), tehát a modellünk statisztikailag szignifikáns. Ebből a táblázatból, az eredmény összegzésénél, szükségünk lesz az *F, df* és *n* értékekre (szintén kiemelve).

Negyedik lépés a *független változók elemzése* a *Coefficients* táblázat alapján. A standardizált *Beta* oszlopban feltüntetett értékek a független változók egyéni hozzájárulásait jelzik a függő változó magyarázásában. A legmagasabb beta értékkel rendelkező változónk lesz a legerősebb prediktor. A *Beta* oszlopban az értékek esetleges negatív előjelét figyelmen kívűl kell hagyni. Továbbá szükségünk van minden független változó esetén a *p* értékre is, mivel csak a statisztikailag szignifikáns ($p<.05$) változóknak lesz egyéni hozzájárulásuk a modellhez. A *Part* részkorrelációs koeficienseket ha négyzetre emeljük (megszorozzuk önmagával, majd a kapott eredményt beszorozzuk 100-al), megkapjuk, hogy a függő változó varianciájának hány százalékát magyarázzák az adott független változók. A *Part* korrelációs együtthatók négyzetre emelt, majd százalékká alakított összeadott értékei nem fognak egyezni az *R Square* értékkel, mivel a részkorrelá-

ciós értékek csak az adott független változó egyedi hozzájárulást jelenti a függő változóhoz, átfedések nélkül.

Coefficients[a]

Model	Unstandard. Coeff.		Stndrd Coeff.	t	Sig.	95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1 (Cnst)	21,587	2,440		8,848	,000	16,790	26,385					
HP	-,062	,004	-,727	-16,047	,000	-,069	-,054	-,661	-,640	-,613	,711	1,406
Model Year	,015	,030	,020	,502	,616	-,043	,073	,249	,026	,019	,878	1,139
Cntr. of Origin	-,530	,142	-,160	-3,724	,000	-,810	-,250	,176	-,190	-,142	,791	1,265

a. Dependent Variable: Time to Accelerate from 0 to 60 mph (sec)

A fenti táblázat kiemelt oszlopaiból kitűnik, hogy a mintában szerepelő autók gyorsulásának magyarázásában a teljesítmény és a gyártási hely is szignifikáns tényező, viszont a legerősebb prediktor a teljesítmény (*beta* = -.727). A gyártási év statisztikailag nem szignifikáns. A teljesítmény 37,95%-ban, a gyártási hely elhanyagolható, mindössze 2,56%-ban magyarázza a gyorsulás, mint függő változó teljes varianciáját (Part korrelációs együtthatók négyzetre emelve, majd százalékka alakítva).

Mint láthatjuk, a didaktikai célzattal készített modell banális, mivel csak egyetlen tényezőnek sikerült kimutatni az érdemi hozzájárulását, mint prediktor, az autók gyorsulásában. A modell számos olyan paramétert figyelmen kívűl hagyott, mint például a hajtáslánc áttételezését (mert nem volt rá adat), a tömeget és fogyasztást (az egyszerűsítés kedvéért nem vontuk őket össze, hanem töröltük), amelyek köztudottan meghatározóak a gyorsulás szempontjából.

Ha nem megfelelően választjuk ki a változókat, vagy ha az adatgyűjtés nem megfelelő, az eredmények hibásak vagy torzítottak lehetnek.

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

Standard többszörös regressziós eljárással vizsgáltuk a mintában szereplő autók gyorsulásának prediktorait. Független változóink a teljesítmény, gyártási év és gyártási hely voltak. Az előzetes elemzés során vizsgáltuk a normalitás, linearitás, multikollinearitás és homoszkedaszticitás feltételeinek teljesülését, amelyeket műveletalkalmassági szempontból megfelelőnek találtunk (utalás a függelékekre, ahol megadjuk a korrelációs táblázatot, normál valószínűségi diagramot és a szórásdiagramot).

A független változókat egyszerre adtuk meg, majd az elemzés elvégézése után megállapítottuk, hogy a modell az autók gyorsulásának, mint függő változó teljes varianciájának 45,7%-át magyarázza (F, (3,375) = 101.414, $p < .01$).

A legerősebb egyéni hozzájárulása a teljesítmény prediktorának van a gyorsulás függő változóhoz viszonyítva (*beta* = –.727, $p < .001$), amely egyben a függő változó teljes varianciájának 37,95%-át magyarázza.

A gyártási hely is statisztikailag szignifikáns prediktor, de ennek magyarázó hatása mindössze 2,56% (*beta* = –.16, $p < .01$).

A gyártási év nem fejt ki szignifikáns egyéni kontribúciót a gyorsulás magyarázásában (*beta* = .02, $p>.05$).

A *Model Summary, az ANOVA* és a *Coefficients* táblázatokat a függelékekhez mellékeljük.

Jelölése: lásd az ANOVA próbánál.

12.7.4. Logisztikus regresszió

A logisztikus regresszió során egy kategorikus változó előrejelzésére alkalmas modellt állítunk fel. A művelet egy szett prediktor hatását vizsgálja a kategórikus változóra nézve. A kategórikus változót függő változónak is nevezzük, mivel a független változók, ún. prediktorok befolyásolják. A kategórikus változó általában egy eseményre utal, a prediktorok az esemény bekövetkezésének valószínűségét jelezhetik előre. Ez a valószínűségszámolás a statisztikában az angol *odds ratio* néven vált ismertté, esélyhányadosnak, vagy esély-aránynak nevezünk; jelölése *OR*. Az esélyhányados segítségével értelmezhetjük, hogy milyen hatással vannak a prediktorok az esemény valószínűségére.

Többféle kiszámítási technikája is ismert: a bináris[56], a multinominális[57] és az ordinális[58]; műveleti szempontból is több lehetőség kínálkozik, például a prediktorok egy blokkban történő tesztelése, vagy lépcsős tesztelése.

Jelen kézikönyvben csak a leggyakrabban használt, a független prediktorok egy blokkban történő tesztelésére térünk ki.

A minta mérete befolyásolja az elemzés statisztikai erejét: a változónkénti tíz esemény, bár széles körben elfogadott minimális kritérium, mégis fenntartásokkal kell kezelni, mivel a gyakorlat nem mindig támasztja alá. A változónkénti esemény alatt nem a teljes minta alanyait kell érteni, hanem csak a kisebbik eredménycsoport alanyait a becsült regressziós együtthatók számához képest. A kis adathalmazokon alapuló regresszív modellek rossz előrejelzésekhez vezetnek, amelyek túl szélsőségesek vagy bizonytalanok (van Smeden et. al., 2016). Nehéz

56 A függő változó dichotóm (bináris), vagyis csak két értéket vehet fel.
57 A függő változó több mint két nominális értékkel rendelkezik.
58 A függő változó ordinális mérési szintű.

pontosan meghatározni a minta méretét, mivel „a logisztikai modell teljesítménye a változónkénti események számán kívül számos más tényezőtől is függhet, beleértve a kovariánsok és az eredmény közötti összefüggések erősségét és a kovariánsok közötti korrelációt" (van Smeden et. al., 2016, 2. o.).

A logisztikus regressziónak kevesebb műveletalkalmassági feltétele van a változókra vonatkozóan, mint az előző fejezetben tárgyalt többszörös lineáris regresszió esetén, azonban ezeket előzetesen, külön műveletekkel szükséges ellenőrizni.

Az analízis egyik feltétele a multikollinearitás ellenőrzése; ideális esetben a függő és független változók közötti korreláció erős (r = .03 - 0.7), a független változók között lehetőleg gyenge. Amennyiben túl erős a korreláció két független változó között (r > .07), akkor csak az egyiküket érdemes bennehagyni a modellben, esetleg össze is lehet vonni őket.

Mivel az SPSS a logisztikus regresszió közvetlen beállításainál nem tartalmazza a változók közötti korrelációk előzetes vizsgálatának lehetőségét, többféle módon is eljárhatunk: előzetesen korrelációs táblázatokat készítűnk a függő és független változóink közötti kapcsolat vizsgálatára (*Analyze → Correlate → Bivariate*) vagy a Pallant (2007) által javasolt, az előző alfejezetben ismertetett többszörös lineáris regresszió során alkalmazott kollinearitás diagnosztika szerint járunk el (*Analyze → Regression → Linear*: *Statistics* fül alatt *Collinearity diagnostics.* A *Coefficients* táblázatból a *Collinearity Statistics* oszlopot kell figyelembe venni, ahol a .1-nél kisbb *Tolerance* értékek a független változók közötti erős korrelációra utalnak).

Az analízis másik feltétele, hogy ne legyenek kiugró értékek, amelyekre érzékeny a modell. Ezeket is előzetesen meg kell vizsgálni, ahogyan az előző alfejezet szerint a többszörös lineáris regresszió esetén is tettük.

Feltételei:

– egy függő változó, amely kategórikus;

– kettő vagy több független változó (prediktor), amelyek lehetnek folytonosak és/vagy kategórikus változók (lehet akár mindkettő egy adatszettben);

– a prediktor változók adateloszlásának nem feltétele a normlitás;

– A predikciós modell értelmezhetőségének érdekében nagyon fontos a változók kódolása (Pallant, 2007): a dichotom változók esetén javasolt a 0/1-es kódolás, ahol „0" az esemény hiányára, az „1"-es a meglétére utaljon; továbbá fontos, hogy a folytonos változóknál a magas értékek a legjellemzőbb tulajdonságokat jelöljék;

– előzetesen ellenőrizni kell a kollinearitást és a kiugró értékeket a reziduumok alapján.

Parancssor és beállítások:

Mivel a **PSPP** műveleti lehetőségei korlátozottak, csak az **SPSS**-re vonatkozóan adjuk meg a beállításokat.

Figyelem: a lent megadott parancssor csak a dichotom (bináris) függő változók esetében érvényes. Amennyiben a függő változónk kettőnél több kategóriával rendelkezik, a *Multinomial Logistic* eljárást kell lefuttatni.

SPSS:

1. Edit → Options be kell jelölni a *No scientific notation for small numbers in tables*.

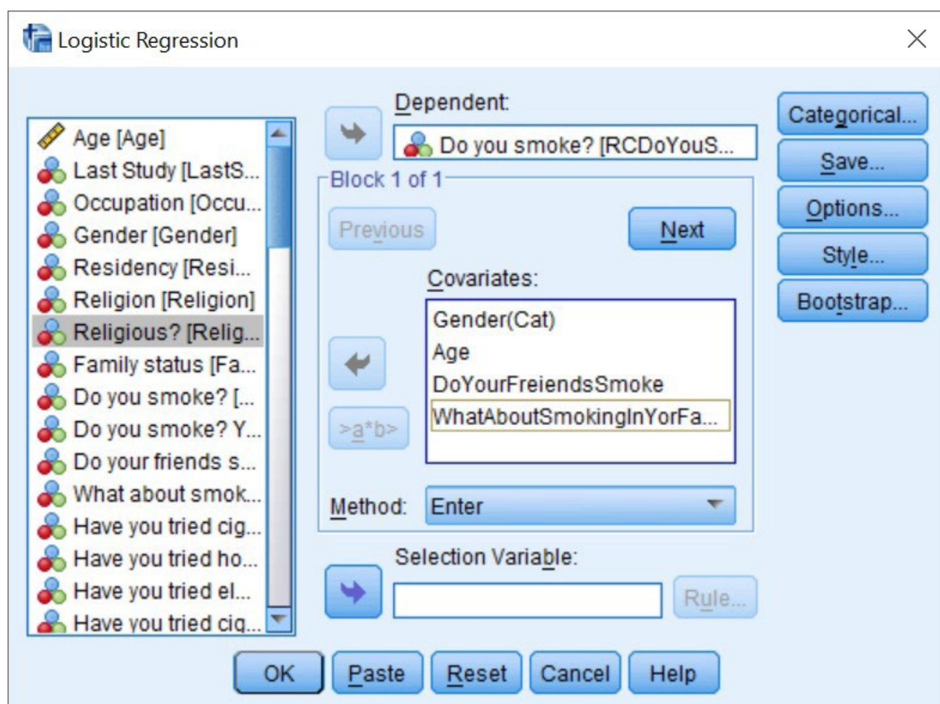
2. Analyze → Regression → Binary Logistic

A *Dependent* mezőbe betesszük a dichotom függő változónkat. A *Covariates* mezőbe betesszük a prediktor változóinkat. Alatta a *Method* típusánál megadjuk az *Enter*-t (a prediktorok egy blokkban történő elemzése). Amennyiben nominális mérési szintű változóink is vannak a prediktorok között, ezeket a *Categorical* fülre kattintva külön betesszük a *Categorical covariates* mezőbe. Ugyanebben az ablakban minden kategórikus változónkat egyenként bejelöljük és a *Change contrast*-ban átállítjuk *First*-re (a változó neve után meg fog jelenni kerek zárójelben, hogy „*first*"). Az Options fül alatt bejelöljük a következőket: *Classification plots, Hosmer-Lemeshow goodness of fit, Casewise listing of residuals* és *CI for Exp(B)*.

Az eredmény értelmezése és magyarázatok egy példán keresztül:

A példa alapjául egy korábbi kutatásunk adatbázisa szolgál, melyben fiatal felnőttek dohányzási szokásait és az ezzel kapcsolatos attitűdjeit mértük fel. Logisztikus regressziós analízissel azon tényezőket (prediktorokat) szerettük volna feltárni, amelyek valamilyen mértékben előrejelezhetik, hogy egy adott személy dohányzik-e vagy sem. A regressziós modell függő változója egy dichotom változó (az alany dohányzik-e vagy nem – igen/nem), a független változók (prediktorok) pedig olyan körülmények, amelyekre a kutatás során rákérdeztünk: nem (bináris változó), életkor (intervallum), dohányoznak-e a barátai (skála) és dohányoznak-e a családban (skála).

A beállításokat a fentiek szerint elvégeztük, megadtuk a változókat. Mivel az alanyok nemére vonatkozó változónk bináris, a *Categorical* fülre kattinva külön betettük a *Categorical covariates* mezőbe, majd a *Change contrast*-ot átállítottuk *First*-re és lefuttatuk az elemzést.



A kimeneti ablakban több táblázatot is kapunk, melyeket négy szekcióba rendezve tekinthetünk meg:

1. Eseti összefoglaló és változók kódolása - az első táblázat az esetek gyakoriságát és százalékos arányát tartalmazza (*Case Processing Summary*), a második táblázat a függő változó kódolását (*Dependent Variable Encoding*).

2. Block 0: Begenning Block – a független változók nélküli eredményeket láthatjuk, melyek valószínűségi becslésen alapulnak, tehát egyfajta viszonyítási alapként tekinthetünk erre az eredménytömbre.

3. Block 1: Method = Enter – a függő változókkal tesztelt modell eredményeit tartalmazza

4. Kiugró értékek – a *Casewise list* táblázatban megjelenik azoknak az eseteknek a sorszáma, amelyek nem illeszkednek a modellbe (2-nél nagyobb standardizált reziduumok).

Az elemzés lépései:

– megvizsgáljuk, hogy az esetek hány százalékát azonosította a modellünk; ehhez összevetjük a *Block 0* és *Block 1*-es *Classification table* utolsó oszlopainak legalsó sorait *(Percentage Correct).* Ha a *Block 1*-ben szereplő százalékos érték nagyobb, mint a *Block 0* százalékos értéke, akkor a független változóinkkal sikerült pontosítanunk a modellt. Példánkban modellünk az esetek 74%-át sorolta be helyesen (*Block 1*), ami nagyobb, mint a *Block 0* 62,8%-os értéke.

Classification Table[a]

			Predicted		
			Do you smoke?		
	Observed		No	Yes	Percentage Correct
Step 1	Do you smoke?	No	73	43	62,9
		Yes	38	158	80,6
	Overall Percentage				74,0

a. The cut value is ,500

– ellenőrizzük a *Block 1*-ben modellünk megbízhatóságát: az *Omnibus Tests of Model Coefficients* táblázat khi-négyzeten alapuló illeszkedésvizsgálata szignifikáns kell, hogy legyen (p<.05).

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	71,310	4	,000
	Block	71,310	4	,000
	Model	71.310	4	,000

– a *Hosmer and Lemeshow Test* is egy nagyon megbízható illeszkedésvizsgálat, amely, ha nem szignifikáns, akkor jó a modell.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	10,009	8	,264

– a *Model Summary* táblázatban megnézzük, hogy a függő változó varianciájának hány százalékát magyarázza a modell, ami jelen esetben 20,4% és 27,9% közé esik.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	340,471[a]	,204	,279

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

– megállapítjuk a prediktorok modellhez viszonyított egyéni hozzájárulását, és beazonosítjuk a fő befolyásoló tényezőt a *Variables in the Equation* táblázat alapján.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Sex	-,111	,298	,138	1	,711	,895	,499	1,607
	Age	-,024	,022	1,267	1	,260	,976	,935	1,018
	DoYourFreiendsSmoke	-1,400	,246	32,430	1	,000	,247	,152	,399
	SmokingInYorFamily	-,602	,140	18,420	1	,000	,548	,416	,721
	Constant	5,786	1,004	33,213	1	,000	325,763		

a. Variable(s) entered on step 1: Sex, Age, DoYourFreiendsSmoke, WhatAboutSmokingInYorFamily.

A táblázatból a kiemelt oszlopok információtartalma fontos: *B, Wald, Sig*. és *Exp(B).*

***B* oszlop:** a függő változónk "igen" válaszának valószínűsége; a +/– előjel a kapcsolat irányát adja meg, a pozitív előjelű független tényezők növekedése esetében a függő változónk *„1-es kódolású igen"* válaszainak valószínűsége is növekedni fog, a negatív előjelű független tényezők növekedése esetén a független változónk *„1-es kódolású igen"* válaszok valószínűsége csökkenni fog.

***Wald* oszlop**: hipotézis teszt, a fő befolyásoló tényezőt azonosítja;

***Sig.* oszlop**: a modell szempontjából szignifikáns prediktorokat jelzi;

***Exp(B)* oszlop:** az esélyhányados *OR* értéke;

95% C.I.for EXP(B) oszlopok: ebben az értéktartományban (*Lower* és *Upper* között) 95%-ban megbízható az esélyhányados valódi értéke.

– Utolsó táblázatunk a kiugró értékeket tartalmazó esetek sorszámait listázza, amelyek nem illeszkednek a modellbe (2-nél nagyobb standardizált reziduumok). Mivel a kiugró értékek a függő változó *„0 – nem"* kategóriájú válaszok prediktoraként kerülnek be a modellbe, csökkentik a modell pontosságát.

Casewise List[b]

Case	Selected Status[a]	Observed	Predicted	Predicted	Temporary Variable	
		Do you smoke?		Group	Resid	ZResid
16	S	N**	,924	Y	-,924	-3,488
75	S	N**	,921	Y	-,921	-3,404
94	S	N**	,917	Y	-,917	-3,322
187	S	N**	,917	Y	-,917	-3,322
280	S	N**	,921	Y	-,921	-3,404
309	S	N**	,919	Y	-,919	-3,363

a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

b. Cases with studentized residuals greater than 2,000 are listed.

Mérlegelnünk kell, hogy a táblázatban jelzett kiugró értékekekkel rendelkező esetekkel hogyan járunk el. Javasolt egyenként megvizsgálni őket, majd eldönteni, hogy benne hagyjuk-e a modellben vagy töröljük őket (például, ha mérési hibáról van szó). Általánosan alkalmazható megoldás nincs. További támpontok érdekében javasolt a szakirodalom alapos áttekintése (például LaValley, 2008).

Az eredmények megjelenítése a dolgozatban (a példa alapján)**:**

Logisztikus regressziós analízissel vizsgáltuk a dohányzást befolyásoló tényezők valószínűségét. A modellben négy független változó szerepelt: *nem*, *életkor*, *dohányoznak-e a barátai*, *dohányoznak-e a családban*.

A prediktorokat tartalmazó teljes modell statisztikailag szignifikáns volt, χ^2 (4, n = 312) = 71,31, p<.01, ami azt jelzi, hogy a modell képes volt megkülönböztetni azokat a válaszadókat, akik dohányoznak, és azokat, akik nem dohányoznak.[59]

A modell egésze a dohányzás, mint függő változó teljes varianciájának 20.4% és 27.9% közötti arányát magyarázza (*Cox és Snell R négyzet*); az esetek 74%-át helyesen sorolta be.[60]

A négy független változóból csak kettő rendelkezik egyedi szignifikáns hozzájárulással a modellhez (dohányoznak-e a barátai, dohányoznak-e a családban). Ezek közül a legerősebb prediktor a családban való dohnyzás, melynek esélyhányadosa *OR* = 5,48, a második prediktor a dohányzó barátok, melynek esélyhányadosa *OR* = 2,47 (az értékeket a *Variables in the Equation* táblázatból kaptuk, a táblázatot mellékelni kell). Ez azt jelenti, hogy azon személyek esetében, akiknek a családjában mások is dohányoznak, 5,5-ször valószínűbb, hogy maguk is dohányozni fognak; a dohányzó barátok is hozzájárulnak ehhez, 2,5-szörös valószínűséggel.

Figyeljünk a *B* érték előjelére, ami fordított viszonyra utal. Példánkban, mivel az adatbázis kódolása fordított volt, a kis értékek azt jelölték, hogy nagyon jellemző a családban, illetve a baráti körben a dohányzás, ezért az eredményt nem értelmezzük fordított viszonyként.)

Jelölése:

$$\mathbf{X^2\ (4,\ \mathit{n} = 312) = 71.31,\ \mathit{p} < .01}$$

ahol Khi-négyzet (szabadsági fokok (*df*), esetek száma (*n*)) = a teszt eredménye, *p* érték

OR = 5,48

ahol *OR* = esélyhányados

59 A khi négyzet értékek a *Block 1*-ből származnak (*Omnibus Tests of Model Coefficients* táblázat), az *N* érték a legelső *Case Processing Summary* táblázatból (*Selected Cases Included in Analysis*).

60 Az adatokat a *Block 1*-est követő két táblázatból származnak: *Model Summary* és *Classification Table*.

13. AZ EREDMÉNYEK BEMUTATÁSA

Az eredmények bemutatására, attól függetlenül, hogy kvantitatív vagy kvalitatív módszereket használtunk, az alábbi megjegyzések egyaránt érvényesek:

A legfontosabb egy megfelelő sorrend vagy struktúra kialakítása, amely mentén az eredmények bemutatásra kerülnek. Ajánlott, hogy a sorrend tematikus jellegű legyen. A bemutatás sorrendje nem az elvégzett statisztikai műveletek sorrendjét követi, hanem a tematikus rendezést.

A dolgozat struktúráját ajánlott úgy kialakítani, hogy az egymáshoz kapcsolódó eredmények egy egységet képezzenek a leírásban. Figyeljünk arra, hogy ne ismételjük a már bemutatott eredményeket, tehát egy adott eredmény a dolgozatban csak egyszer szerepeljen, annak minden összefüggésével együtt.

A struktúra kialakítását valamilyen általunk választott logikai szempont kell vezérelje. Minden egyes kutatás egyéni megközelítést igényel, erre nincs általános recept.

Az eredményeinket vissza kell csatolni az elméleti részhez, ez adja meg a szakdolgozat egységességét. Ez abból áll, hogy a kapott eredményeinket összevetjük a szakirodalomban felleltekkel, és megállapítjuk, hogy azzal megegyezőek-e vagy attól eltérőek. Eltérő eredmények esetén próbáljunk magyarázatot találni a jelenségre. Például, ha az online zaklatás jelenségét vizsgáljuk öt éves távlatban, a friss adataink feltehetőleg nagyobb gyakoriságot fognak mutatni, mint a korábbiak. Ennek lehetséges magyarázata az, hogy a közösségi oldalak használata idővel jobban elterjedt.

A visszacsatolásnál használjuk a szakirodalmi hivatkozásokat, de korlátozódjunk az eredményekre, ne részletezzük az elméleti hátteret.

Sokat segíthet a dolgozat összeállításában, ha az elméleti rész elkészítése után, és az tanulmányokban közölt kutatási eredmények ismeretében szerkesztjük meg a kérdőívet, így minden várható eredményünk összehasonlítható lesz más vizsgálati eredményekkel.

Amennyiben kvantitatív adatfeldolgozást végeztünk, figyeljünk adataink közlési formátumára is. Amennyiben egy statisztikai mutató értéke 0-val kezdődik, csak a tizedes értékeket írjuk le, kerekítve és két tizedes pontossággal. Például ha az általunk számolt szignifikancia szint $p = 0.0000235$, akkor a következő jelölést használjuk: $p < .01$. Ugyanígy ha egy korrelációs együtthatónk $r_s = 0.325$, akkor így jelöljük: $r_s = .33$.

14. A DOLGOZAT ÖSSZEFOGLALÁSA ÉS JAVASLATOK MEGFOGALMAZÁSA

Minden tanulmányt és szakdolgozatot egy összefoglaló fejezettel zárunk, melyben röviden összegezzük az eredményeinket, valamint megfogalmazhatjuk a témával kapcsolatos javaslatainkat. Az összegzés terjedelmének arányulnia kell a dolgozat terjedelméhez, ami egy 30-50 oldalas kutatás esetén 1-2 oldal szokott lenni. Többféle módon is elvégezhetjük az összegzést, a kutató habitusa, tapasztalata, késztetése szerint. Törekednünk kell arra, hogy tömör nyelvezettel ragadjuk meg a főbb tendenciákat. Én két, gyakran alkalmazott megoldást javasolnék:

A legjárhatóbb út az, hogy kövessük a dolgozat struktúráját: utalunk a kutatás célkitűzésére, ennek teljes vagy részleges megvalósulására, röviden ismertetjük a szakirodalom aktuális álláspontját a témában, újrafogalmazzuk az alkalmazott módszereket és a fontosabb eredményeinket.

A szakirodalom álláspontjának összegzésekor továbbra is következetesen meg kell jelölnünk a hivatkozásokat, viszont javasolt összevontan feltüntetni a szerzőket. Például leírunk egy tendenciát, amire a témában végzett több korábbi tanulmány is utalt, a szerzőket és évszámokat egymás után tesszük kerek zárójelbe: (Szerző1, évszám; Szerző2, évszám; Szerző3, évszám). Mivel az összegzésben nem szoktunk idézeteket bemásolni, az oldalszámokat nem jelöljük meg. Olyan szakirodalom, amit a dolgozat korábbi részében nem használtunk, az összegzésben nem szerepelhet, ebből következik az is, hogy nem hozhatunk fel olyan új tartalmi elemeket vagy összefüggéseket, amelyeket korábban nem elemeztünk.

Az eredmények összefoglalása előtt pár mondatban a kutatásunk módszertanáról is szükséges szót ejteni. Konkrét számadatokat nem szükséges ismételten leírnunk, elegendő a megfigyelt tendenciák alapján megragadni a lényegi következtetéseinket. Az összegzésben semmiképp sem javasolt részletekbe bonyolodni.

A dolgozat egységessége szempontjábói előnyös, ha az összegzésben is megemlítjük, hogy a saját eredményeink hogyan viszonyulnak a szakirodalomban jelzettekkel, azonos vagy eltérő következtetésekre jutottunk-e.

Egy másik összegzési javaslat, mely szintén könnyen járható: az összegzést a célkitűzés felemlítésével kezdjük, rögtön rátérve a kutatás módszertanára. Következik a fontosabb eredményeink tendencia szintű összefoglalása, és ezzel párhuzamosan a szakirodalom aktuális álláspontjának az összefoglalása,

nagyon röviden. Ez az összegzési struktúra abban tér el az elsőtől, hogy nem térünk ki külön a szakirodalom álláspontjára, az eredményeink bemutatásával párhuzamosan ismertetjük azt.

Az összegző fejezetet zárhatjuk a kutatás során megfogalmazott javaslatainkkal. Ezek a javaslatok vonatkozhatnak például az adott téma jogszabályi hátterére, a szakmai jógyakorlatra, a téma iránti társadalmi érzékenyítésre, vagy bármilyen általunk fontosnak vélt szempontra. Javasolt a kutatás gyakorlati hasznát is megfogalmazni, mivel ez a védés során az értékelő bizottság részéről gyakran felmerülő kérdés.

15. A TARTALOM VIZUALIZÁCIÓS ESZKÖZEI

A dolgozat minősége fokozható adekvát sematikus tartalommegjelenítéssel. Amikor megtervezzük a kutatást, lépésről lépésre haladunk előre, egy bizonyos struktúrát, szempontrendszert követve. Rendszerint a szempontrendszert a célkitűzéshez szabjuk. A különböző tematikus egységeket fejezetekbe és alfejezetekbe rendezzük, egy logikai fonal mentén. A kutatás lépéseit, folyamatát különböző ábrákban és összegző táblázatokban is megjeleníthetjük. Ennek előnye, hogy segíti a tartalom befogadását az olvasó, vagy értékelő számára.

Az alábbiakban néhány példát említenék, hogy milyen eszköztárból meríthetünk a tartalom szemléletessé tételéhez:

– A tartalommegjelenítésre használhatunk **folyamatábrákat**, amelyeket először akár papíron is megtervezhetünk, majd a *Word* szövegszerkesztő *SmartArt* funkciója segítségével digitalizálhatunk. A folyamatábrákban kulcsszavakkal jelölhetjük a tematikus egységeket; a köztük lévő összefüggéseket, kölcsönhatásokat vagy oksági összefüggéseket, megfelelő irányú nyilakkal jelölhetjük. Amennyiben két vagy több tematikus egység között átfedés van, használhatunk különböző geometrikus alakzatokat is, amelyek részleges vagy teljes átfedésben vannak.

– **Összefoglaló táblázatok** alkalmazása. Az elméleti részben, például, az olvasott szakirodalmat olyan táblázatokban is összefoglalhatjuk, amelyek tartalmazzák a szerzőt, címet, publikálás évét, és 1-2 tömör mondatban az adott tanulmány következtetéseit, eredményeit. Ez a fajta összegzés hasonlít a metaanalízis módszeréhez, mely során statisztikai elemzéssel vetjük össze több tudományos mű vagy tanulmány eredményeit, azzal a különbséggel, hogy nem végzünk semmilyen statisztikai elemzést. Dolgozatokban gyakran előfordul, hogy a hallgató az összesítő táblázatra metaanalitikus táblázatként hivatkozik, ami hiba, ha nincs mögötte statisztikai elemzés. Célszerűbb a szakirodalmi összesítés vagy más hasonló megnevezést használni.

– A tartalom megjelenítésére használhatunk **SWOT elemzést**, mely egy adott szempontrendszer szerint sorolja be az eredményeinket egy 2x2-es matrixba. A szempontrendszer négy eleme az erősségek (*strenghts*), gyengeségek (*weaknesses*), lehetőségek (*opportunities*) és veszélyek (*threats*):

SWOT elemzés

	hasznos	ártalmas
Belső tényezők	erősségek (Strenghts)	gyengeségek (Weaknesses)
Külső tényezők	legetőségek (Opportunities)	veszélyek (Threats)

– **Létező modellek adaptációja:** Amennyiben korábbi ismereteink vagy olvásmányaink alapján állítunk fel modelleket, a forrás megjelölésével bátran válogathatunk ezek közül is, sőt, akár tovább is fejleszthetjük őket. Példaként említeném akár a pszichológiából ismert *Maslow* féle szükséglet piramist, a *Shramm* féle kommunikációs modellt, a *Hart* féle döntéshozatalban való részvételi létra modelljét. Tulajdonképpen bármilyen tudományos modellt felhasználhatunk. A modellek közös jellemzője, hogy adott témában megragadnak bizonyos jellemzőket és ezeket egy bizonyos szempont szerint jelenítik meg. Többféle modell létezik, például: analógiára épülő modellek, amelyek jelenségeket vagy tulajdonságokat állítanak párhuzamba, léteznek idealizált modellek, amelyek az egyszerűsítésre törekszenek, vagy magyarázó modellek, amelyekkel különböző elméletek sajátosságait magyarázhatjuk.

A tartalom különböző formákban történő megjelenítését, vizualizációját a dolgozat bármely részében használhatjuk, legyen szó elméletről, módszertanról vagy eredmények bemutatásáról. Egy kivétel azonban akad: a szakdolgozatot záró, összegző fejezetben már csak szöveges formában foglalhatjuk össze a kutatás lényegét, mivel minden egyes vizualizációs eszköz alkalmazása rendszerint egy új szemponton alapszik.

16. A SZAKDOLGOZAT VÉDÉSE ÉS A PREZENTÁCIÓ

A védés véleményem szerit egy vizsgával összekötött ünnepélyes alkalom, amely során a hallgató saját magának, a tanárainak és végzős társainak bizonyíthatja, hogy kiérdemelte az egyetemi oklevelet. A védés kapcsán is felmerülhetnek kisebb-nagyobb botlasztók, melyekre szeretném felhívni az olvasó figyelmét, valamint néhány javaslatot is megfogalmaznék.

A védés időtartama rendszerint 10+5 perc, amiből az első 10 perc alatt a hallgató prezentálja a szakdolgozatát, majd a vizsgabizottság teszi fel kérdéseit a következő 5 percben. Tapasztalataim szerint ebbe az időkeretbe percenkét egy-másfél dia (slide) férhet bele, tehát a 10-15-nél több diát tartalmazó bemutatókkal túlléphetjük az időkeretet. Javasolt otthon előzetesen begyakorolni a bemutatást, hogy a hallgató ne lépje túl a rendelkezésre álló időt.

A prezentáció a dolgozat bemutatását szolgálja, de nem mindegy, hogy hogyan építjük azt fel. Tükröznie kell a dolgozat szerkezetét, de nem feltétlenül kell annak terjedelmi arányai követnie: a prezentációban kevesebb az elmélet bemutatására eső rész (javasolt mindössze 2-3 dia - slide), a nagyobb rész az eredmények bemutatására összpontosítson. A prezentációs oldalaknak áttekinthetőeknek kell lenniük, így nem érdemes túlzsúfolni őket; a betűméretet se aprózzuk túlságosan el, mert a bemutatóterem ülőhelyeinek mindegyikéről jól olvashatónak kell lennie. A háttér megválasztása mellékes szempont, de a kontrasztra érdemes figyelni, tehát vagy világos alapon sötét betűket, vagy sötét alapon világos betűket használjunk. Az ábráknak és táblázatoknak helye van a prezentációban (lehetőleg tartsuk be itt is az APA hivatkozási stílus követelményeit). Bizonyos kulcsszavakat kiemelhetünk, használhatunk dőlt vagy félkövér betűtípust, de a diákat túlzottan ne színezzük ki. A színes ábrák elfogadhatóak, sőt javasoltak, de a szöveg ne legyen színes.

A szöveges részben a teljes mondatok helyett használjunk kulcsszavakat és vázlatszerű összefoglalókat, ezek emlékeztető mankók, amelyekre vizsgahelyzetben szükségünk lehet. Akár papír alapú jegyzeteket is magunkkal vihetünk, viszont tartózkodjuk a bemutató felolvasásától. Ha még nem kiforrott a szabatos előadásmódunk, akkor is jobb ezzel próbálkozni, mert a felolvasással monotonná válik a prezentációnk, amit a bizottság nem fog értékelni. Nem ritka, hogy a bizottság jelzi a hallgatónak, hogy beszéljen szabadon. Védéskor minden résztvevő és elbíráló tisztában van a hallgatók stressznek való kitettségével, ezért az előadásmód tekintetében általában elnézőek és megértőek.

Az elmúlt 15 év oktatói tapasztalatom szerint jellemző hibák a következők:

1. Egy hallgató a védés előtt pár perccel elveszítette az eszméletét. Szerencsére szinte azonnal megtaláltuk a mosdóban és értesítettük a mentőket, így nem lett komolyabb baj. Kiderült, hogy a védés előtti napokban a stressz miatt nem evett rendszeresen és nem pihente ki magát kellőképpen, ami hozzájárult a pillanatnyi rosszulléthez. Az eset végül még aznap egy sikeres védéssel zárult.

2. Szóbeli vizsgán a tétel kihúzását követően egy hallgató teljesen lefagyott. Lehetőséget kapott egy másik tétel kihúzására, de ez sem segített. Feltehetően itt is a túlaggódás és a túlzott stressz játszott közre, valamint a felkészülés sem lehetett megfelelő. A bizottság türelmesen állt hozzá, sőt még segítő kérdéseket is feltett, próbálta ösztönözni a hallgatót. Külön gondolkodási időt is kapott, hogy próbáljon jegyzeteket vagy kulcsszavakat leírni, de ez sem segített. A vizsga sajnos számára nem zárult sikeresen.

3. Pillanatnyi emlékezészavar, lapszus gyakran előfordul; a hallgatónak csak egy adott szó nem jut eszébe, vagy annak csak a kezdőbetűjére emlékszik. Ilyenkor általában a bizottság segíteni szokott.

4. Többször is előfordult, hogy hallgatók azonosultak az intézménnyel, ahol a kutatást végezték. Ez sajnos az objektivitás elvesztéséhez vezet, ami a kutató egyik legfőbb kvalitása kell, hogy legyen. Több olyan szakdolgozat bemutatón is részt vettem, ahol a hallgatók semmilyen intézményi problémáról nem számoltak be, mintha promotálnák azt.

5. Más esetekben a hallgatók a védés alatt kiálltak, állást foglaltak, személyes véleményt fogalmaztak meg egy bizonyos célcsoport mellett. Ezekben az esetekben is a kutatói objektivitás válik kétségessé, tehát a sem a dolgozatban, sem a védésen nem javasolt jogvédő aktivistaként megnyilvánulni.

6. A hallgató magabiztosan kijelentette, hogy témájának nincs hazai szakirodalma. Éppen a bemutató témában írtam a doktori értekezésemet és néhány szakcikket is publikáltam, tehát a hallgató felületesen dokumentálódott. Az eset tanulsága az, hogy nem árt előzetesen tájékozódni a bizottsági tagok tudományos munkásságáról, megnézni azok szakterületeit.

7. A hallgató magabiztossága a védés során előnyt jelent, mert azt sugallja, hogy alaposan felkészült. Ennek egyik módja a bizottsági tagokkal való szemkontaktus keresése. Viszont olyan védésen is résztvettem, amikor a hallgató kiválasztott magának egyetlen bizottsági tagot és csak rá összpontosított, neki szólóan tartotta a bemutatót. Inkább tekintsük a bizottságot közönségnek, mert mindegyik tagja egyforma súllyal értékel, ezen túlmenően, több személytől is kaphatunk mimikai vagy testbeszédi–visszajelzéseket, amelyek segíthetnek a vizsgahelyzetben.

8. Végül, amint már korábban is említettem, tartózkodjunk a bemutató felolvasásától, ezt a vizsgabizottság nagy valószínűséggel nem fogadja el.

17. A TÁRSADALOMTUDOMÁNYI KUTATÁSOK ÚJ PERSPEKTÍVÁJA

Akik az oktatással bármilyen kapcsolatba kerültek az elmúlt években, metapasztalhatták, hogy jelentős teret nyert a digitalizáció és az ITK eszközök használata is egyre jobban elterjedt. Ezen változásokat a Covid-19 online oktatásra való szükségszerű áttérése felgyorsította (Malatyinszki, 2020). Mivel az oktatás és a kutatás két olyan terület, mely számos kapcsolódási pontot mutat egymással, a digitális technológiák kutatások módszertanára is hatással vannak. A 8. fejezetben már utaltam a mesterséges intelligencia és a nagy nyelvi modellek nyilvános elérhetőségének jelentőségére. Kézikönyvemet a jövőben várható változásokkal zárnám egy frissen közölt tanulmányomra alapozva, átvéve annak szövegét. A tanulmányomban a Big Data alapú elemzések perspektíváit mutattam be a társadalomtudományi kutatásokban (Albert-Lőrincz, 2022).

A társadalomtudományi kutatások során az egyéneket bevonó hagyományos kvantitatív adatgyűjtés több problémát is felvet; bizonyos korlátok közé szorítja az eredmények értelmezhetőségét.

Hagyományos adatgyűjtés alatt azt értjük, hogy a kutató igyekszik kialakítani egy reprezentatív mintát, majd lekérdezi az alanyokat, vagy már meglévő adatbázisokkal másodlagos adatelemzést végez. Mindkét eljárás lényege ugyanaz, a válaszok az egyének által felvállalt többé-kevésbé szubjektív véleményét fogják tükrözni.

Az egyik felmerülő probléma az, hogy a mintavételi eljárás valóban reprezentatív-e, a másik, hogy a felvállalt válaszok mennyire őszinték. Eddigi kutatómunkám során többször szembesültem olyan helyzettel, hogy a kutatásba bevont alanyaink az adatfelvétel során, az anonimitás garantálása mellett sem vállalták véleményüket, amennyiben az, többé-kevésbé, eltért a társadalmi elvárásoktól. Ez gyakran olyankor is fennállt, ha kritikát kellett megfogalmazniuk munkaadójukkal szemben. Amint látjuk, a kvantitatív adatelemzés egyik fő korlátja az adatok megbízhatósága.

Az egyének vélekedése és viselkedése között diszkrepancia tapasztalható, sőt önmagukkal sem őszinték (Stephenz-Davidowitz, 2019). A szerző rámutat arra, hogy online környezetben, főként az internetes keresés során, az egyén akaratlanul is felfedi a valós érdeklődési köreit, értékrendjét, preferenciáit és különféle dilemmáit. Sokat elárulhatnak valakiről keresési szokásai, a kulcsszszavak, amiket használ, az egyes honlapokon eltöltött idő, az online rendelései, nem is beszélve a szociális médiában folytatott tevékenységéről. A fel-

használói szokások a célzott marketinget hatékonyabbá teszik és a felhasználó kereskedelmi profilozásához is alkalmazhatóak. Mivel értékesíthető adatokról van szó, ezeket az online szolgáltatók eltárolják, és a *Big Data* egy kis szeletét fogják alkotni. Mellesleg ebben a helyzetben az Európai Adatvédelmi Rendelet hatástalan (Zarsky 2017; Tesfay et al 2018; McDermott 2017), de erre még a későbbiekben visszatérünk.

Adott tehát a kérdés, hogy a kereskedelmi profil megalkotására alkalmas adatok milyen egyéb területeken hasznosíthatóak és hozzá tudnak-e járulni az egyéni és társadalmi viselkedés kutatásához, megértéséhez. A kérdés megválaszolásához nézzük meg a Big Data fogalomkörét. Bár jelenleg még nincs általánosan elfogadott meghatározása, a fogalom az 1990-es évek elején jelent meg az informatikai forradalommal összefüggésben. A lényege abban rejlik, hogy „eddig nem látott nagyságrendű adatok tárolását, feldolgozását, a rejtett és váratlan összefüggések megtalálását feltételezi", segítségével „olyan következtetéseket lehet levonni, melyek kisebb mennyiségű adat feldolgozása során nem tűnnének fel" (Szűcs & Yoo 2016, 9-10 o.). A szakirodalom egyetért a Big Data legfőbb jellemzőit illetőn, amelyet az 5V felosztásként ismerünk: *volume, velocity, variety, veracity és value* – mennyiség, sebesség, változatosság, valódiság és érték (Naqshbandi 2021).

Kétségtelen, hogy a Big Data új perspektívákat nyitott számos kutatási terület számára. Ezen adatok a társadalmi, politikai és gazdasági tendenciák kiértékelésére is alkalmasak (McNeely & Hahm 2014).

A Big Data lehetőségeinek kiaknázása többek között hozzájárult a Google sikeréhez[61], mivel a cég egyben úttörője is az adattechnológiának. Azok a szektorok, ahol rendelkezésre állt a megfelelő pénz és szaktudás, a Big Data kezdetei óta profitálnak az adatok nyújtotta lehetőségekből: a piackutatás, a marketing, a pénzügy, a befektetési alapok, stb., vagyis egyszóval a versenyszféra terén.

Choi és Hal a Google kereső kulcsszó lekérdezések kapcsán 2009-ben publikáltak egy rövid tanulmányt, melyben bizonyították, hogy egy adott iparágban, a lekérdezések korrelálnak a gazdasági aktivitás jelenlegi szintjével, és így a későbbi adatközlések előrejelzésében segítségünkre lehetnek. 2011-ben Hal, aki a Google vezető közgazdásza, egy újabb tanulmányban azt vizsgálta a kulcsszó lekérések kapcsán, hogy hogyan kaphatnak betekintést az emberek érdeklődési körébe, szándékaiba és jövőbeli cselekedeteibe – ami már túlmutat a pusztán gazdasági hasznosíthatóságon.

Megjegyezném, hogy a felhasználók által használt kulcsszavakkal a Big Data-nak csak egy parányi részét lehet elérni, bár itt, a 2004-es évvel kezdődően, temérdek információhoz lehet hozzáférni. 2012-ben Lohr arról számolt be, hogy a Big Data korában élünk, és ez a következő évtizedben a társadalom minden területét forradalmasítani fogja, minden tudományág egyre inkább

61 A statista.com szerint 2022 júliusában majdnem 84%-os a keresőmotork terén a világpiaci részesedése.

adatintenzívebb lesz. Ovadia már 2013-ban felhívja a figyelmet arra a szerepre, amelyet a Big Data a társadalomtudományokban tölthet be. Hesse, Moser és Riley (2015) a Big Data tudássá való átalakításáról értekezett. Stephenz-Davidowitz (2019), a Google volt adattudósa és a New York Times rovatvezetője még ennél is tovább ment, azt példázva, hogy amit az emberekről eddig gondoltunk, az nagyrészt téves, majd könyvében a Big Data segítségével végzett számos kutatásáról számol be.

Az elmúlt években egyre több publikáció jelent meg a Big Data különböző területeken történő hasznosításáról, beleértve a gazdaságtant, az orvostudomány, közegészségügyet, szociológiát, pszichológiát, tehát az alcímben megfogalmazott kérdésre egyértelműen az a válasz, hogy mindenki, pontosabban minden tudományág profitálhat a Big Data nyújtotta lehetőségekből.

Az alcím második tézise, hogy pontosan mit forradalmasít a Big Data-n alapuló adatelemzés, nem teljesen egyértelmű. A versenyszférában ismert, rendszeresen alkalmazott adatelemzési technikákról van szó, évtizedes tapasztalattal rendelkező elemzők dolgoznak a tőkeerős, innovatív vállalatoknak. Az általuk alkalmazott elemzési módszertan és az algoritmusaik viszont üzleti titoknak minősülnek. Tehát az adatok forradalma jelenleg a versenyhelyzet szülte folyamatos innovációt jelenti.

A társadalomtudományok területén azonban más a helyzet, bár léteznek Big Data elemzésen alapuló kutatások, ezek száma viszonylag alacsony a hagyományosnak tekintett módszerekkel végzett kutatásokkal szemben. Ennek valószínűsíthető okai, hogy ezen területek finanszírozottsága sokkal alacsonyabb a versenyszféráéhoz képest, a humántudományi szakemberek matematikai és informatikai kompetenciái számukra többnyire terület-idegenek, és ezekhez még hozzáadódik a kidolgozott és elfogadott módszertani sablonok hiánya.

A társadalomtudományok – vélhetően – lépéshátrányban vannak a versenyszférához képest az innováció tekintetében. Bár a humántudományok számszerűsítése már a 19. század közepétől növekvő jelentőséggel bír (Runkel 1962; Desrosières 1996), ez a folyamat a személyi számítógépek, internet és elérhető, felhasználóbarát statisztikai programok elterjedésével tovább fokozódott. Annak ellenére, hogy a neves szaklapok és a doktori iskolák fokozottan elvárják az eredmények kvantifikációját, mégsem számítunk adatforradalomra a humántudományok területén.

Az intenzív finanszírozás és a több egymástól távolálló területen is képzett szakemberek viszonylagos hiánya mellett további érvként felhoznánk a Big Data elemzések korlátait, melyekből kifolyólag nem váltják ki a hagyományos, bevált módszereket. Az adatelemzések vitathatatlan előnye, hogy új perspektívát nyitnak és lehetővé teszik a hipotézisek gyors tesztelését, és olyan kutatásokat, amikre eddig nem volt mód. Arra számítunk, hogy a Big Data lehetővé teszi az eredmények megbízhatóságának növelését, pontosabbá tételét.

A Big Data elemzés (angolul „*Big Data analytics*” vagy rövidítve „BDA”) egy olyan komplex folyamat, mely lehetővé teszi a rejtett információk feltárását, mint például rejtett minták, összefüggések, piaci trendek, felhasználói preferenciák. Ezen információk kinyerése alkalmazásokon keresztül történik, melyek prediktív modelleket, statisztikai algoritmusokat és az analitikai rendszerek által működtetett „mi lenne, ha” elemzéseket eredményeznek. Az adatokat ún. adattárházakban (*data warehouse*) vagy adattengerekben (*data lake*) gyűjtik és tárolják. Ezt követi az adatok letisztítása különböző szkriptkészítő eszközökkel vagy adatminőség-javító szoftverek segítségével: kiszűrik az esetleges hibákat vagy inkonzisztenciákat, mint például az ismétlődések, formázási hibák, és végül rendszerezik az adatokat. Az utolsó lépés az adatok szoftveres elemzése adatbányászattal, prediktív elemzéssel, adatmegjelenítéssel, machine learninggel stb. (Chai et al s.a.). A fenti folyamat a kutató vagy adatelemző részéről matematikai, statisztikai és informatikai ismereteket feltételez, tehát az egyéni kutató helyett a több szakemberből álló kutatói csoportok (*team*) a hatékonyabbak.

Létezik egy olyan szűk szelete a Big Data-nak, amit, minimális informatikai és statisztikai ismeretekkel, az egyéni kutató is hasznosítani tud. Ez a módszer nem más, mint a Google kulcsszó lekérdezések elemzése, egy olyan módszer, amit elsősorban a weboldal keresőoptimalizálással (*search engine optimization - SEO*) foglalkozó szakemberek ~~is~~ alkalmaznak. Nielsen (2013) kissé sarkított meglátása szerint a Google kulcsszó lekérdezője nem más, mint az emberi szándékok adatbázisa, amely egy sokkal nagyobb vízió kis része: egy adatbázis, amely a világ összes tudását tartalmazza. A szerző szerint kétségtelen, hogy ezen adatok nagyban hozzájárulhatnak a világ megértéséhez.

A módszer lényege, hogy a *Google Trends* felületen (trends.google.com) lehívható, hogy egy tetszőleges kulcsszóra, vagy kifejezésre milyen gyakran kerestek rá. A felület a keresések relatív számát adja meg, vagyis azt, hogy a teljes keresési mennyiséghez viszonyítva egy adott időszakon belül milyen gyakran kerestek rá adott kulcsszóra. A beállítások között kiválasztható egy adott régió, ország, település, a keresési periódus (2004-től az elmúlt óráig), kategória és internetes kereső. Lehetőség van két vagy több kulcsszó összehasonlítására. Az így kapott információ ingyenes, viszont önmagában még nem elegendő a kutatáshoz, mindössze kiindulási alapnak használható. A ritkábban előforduló kulcsszavakra például nagyon alacsony értékeket ad meg, ami önmagában nem bír különösebb jelentőséggel. Viszont, ha vannak hirtelen kiugró keresési csúcsok, ezeket összevethetjük a közelmúlt eseményeivel, és így már támpontként szolgálhatnak a további kutatáshoz. A mélyítéshez szükségünk lesz az abszolút keresési mennyiségre is, ez az adat azonban nem ingyenes. Ezt a *Google Ads* (https://ads.google.com/) segítségével tudhatjuk meg, de számtalan „third party” alkalmazás vagy extenzió létezik (pl. Keyword Everywhere), amelyek segítségével – szintén fizetés ellenében – hozzáférhetünk a számunkra releváns

adatokhoz. Stephenz-Davidowitz (2019) könyvében számos példát hoz fel arra, hogy az emberek szándékai, preferenciái hogyan mutathatóak ki ezzel a módszerrel, illetve kutatásai során milyen korlátokba ütközött.

Az „adattulajdonlás" kifejezés az elmúlt néhány évben felkeltette a kutatók figyelmét. A kérdés azért releváns, mert meghatározza, hogy ki és milyen mélységig férhet hozzá, vizsgálhatja, kutathatja a Big Data adatokat, tehát milyen mértékben nyilvánosak ezek.

A tulajdonlást két szempontból vizsgáljuk: az egyén szempontjából és vállalati szempontból. A két szempont közötti kapocs az, hogy a Big Data egyének, felhasználók millióinak online tevékenységéből, viselkedési mintázataiból és okos eszközeik által képzett adataiból épül fel, amelyeket nevezhetünk online attribútumoknak. Amint megosztjuk a különböző online szolgáltatókkal, vállalatokkal, eltárolják ezen adatokat és saját céljaik szerint hasznosítják.

Egyéni szempontból – Al-Khouri (2012) szerint – a saját adatok a szó szoros értelmében csak a személyes online tevékenységhez kapcsolódó attribútumokat tartalmazzák. Ezek olyan információk, amelyek a megosztás pillanatáig sajátnak tekinthetők, viszont amikor kifejezett szándékkal vagy implicit módón megosztjuk őket, a tulajdonjogot is átruházzuk a szolgáltatóra. Minden egyes megosztással az adatok tulajdonosainak a száma nő és új adatok generálódnak, új tulajdonjogokat létrehozva. Például, amikor egy bejelentkezett Google fiók felhasználó a weben keresést indít, felhasználói keresési minták generálódnak. Noha a keresési információk nem tartoznak a Google-hoz, a keresési mintákban gyűjtött adatok már igen. Ez egy olyan probléma, amely túlmutat a kereskedelem, az etika és az erkölcs határain, és adatvédelmi problémákhoz, a magánélet védelméhez vezet. A jogalkotónak elkerülhetetlenül újra kell terveznie az adatvédelmi törvényeket nemzeti és nemzetközi szinten (Al-Khouri 2012). A Big Data tehát személyes adatokat is tartalmaz, ezek nem megfelelő kezelése sérelmet okozhatnak, ezért a kutatók felelőssége az adatok legális és etikus felhasználása (Wiltshire & Alvanides 2022).

A normatív beavatkozás szükségességére Al-Khouri már 2012-ben felhívta a figyelmet, erre válaszként az Európai Unió 2016-ban kidolgozta új adatvédelmi rendeletét (Európai Parlament és Tanács 2016), mely 2018 május 25-én lépett hatályba. A személyes adatokkal kapcsolatos egyéni kontroll megerősítése volt az uniós jogalkotó egyik fő célja (van Ooijen 2019), viszont a gyakorlatban az online szolgáltatók és vállalatok a tényleges adatvédelem helyett inkább áttértek az *„agree or quit"* (Houser & Voss 2018 104 o.) modellre, mely szerint a felhasználók vagy elfogadják – megfelelő tájékoztatás mellett – a saját üzleti és adatvédelmi feltételeiket, vagy lemondanak az online szolgáltatás vagy platform használatáról. Megjegyeznénk, hogy a lemondásra nincs alternatíva, mivel jellemzően ezen online vállalatok monopol helyzetben vannak, így az adatgyűjtés és értékesítés továbbra sem áll le, csak a felhasználók „rákényszerült" beleegyezésével törté-

nik. A jelenlegi adatvédelem nem más, mint egy kattintással elfogadott megállapodás, ahol a felhasználónak nincs érdemi mozgástere, „take-it-or-leave-it" alapon működik (Compagnucci 2020).

Vállalati szempontból a kiindulási pont az, hogy kik gyűjtik és tárolják az adatokat, mivel mindkét folyamat költségigényes. Innen következik, hogy egy hatékony adatinfrastruktúra csak néhány nagy profitorientált vállalatnál áll rendelkezésre, ún. domináns technológiai platformoknak, mint amilyen például a Google és a Facebook (Nielsen 2013).

Kutathatóság szempontjából az volna az ideális, ha az adathalmaz a nyilvánosság számára is elérhető lenne. Nielsen (2013) amellett érvel, hogy előnyős lenne, ha az adathalmazt vagy infrastruktúrát non-profit szervezetek vagy esetleg magánszemélyek „laza" hálózata működtetnék, mint a felfedezés és innováció platformját. Azonban, egy ilyen infrastruktúra egyetlen darabja jelentős tőkeköltséggel járna, mivel egy nagy számítógépcsoport üzemeltetését feltételezi. A valóság az, hogy a fent említett domináns technológiai platformok magántulajdonban vannak, a platform tulajdonosa pedig megválaszthatja, hogy kivel és milyen mélységig osztja meg az adatokat, hogy versenyelőnyhöz jusson, az emberi tudás feltérképezése által. Nielsen (2013) eszmefuttatásában kifejezi azon reményét, hogy a jövőben a legjobb adatinfrastruktúra bárki számára elérhető lesz a világon egy olyan hatékony platformként, mely lehetőséget biztosít a kísérletezéshez és felfedezéshez.

Bár Nielsen megállapításait közel egy évtizede fogalmazta meg, ami az információs technológiák világában hosszú idő, ma is érvényesek. Azonban, az adatok tulajdonlásának kérdése ma sem tekinthető lezártnak. Fadler és Legner (2020) szerint újra kell értelmeznünk az eddigi koncepciókat, hogy megfeleltessük a Big Data elemzések (BDA) során felmerülő kihívásoknak. Összehasonlító tanulmányukban három adattulajdonlási típust azonosítanak, amelyek a gyakorlatban is előfordulnak: adat, adatplatform és adattermék. Szerintük a jogok mellett a felelősségek is tisztázásra szorulnak.

A fenti felosztásra épített analógia szerint, ha Big Data elemzést végez egy kutató, maga az adat és a felhasznált adatplatform nyilván nem a sajátja. Viszont, ha rendelkezik hozzáférési jogosultsággal, az elemzést követő adattermék, rendszerint tanulmány, már a saját szellemi tulajdonát fogja képezni. Az adatok forrását és a platformot azonban meg kell neveznie, az esetleges személyes adatokat el kell fednie és az általános kutatói etikát be kell tartania.

A nagy áttörés, amelyet az adatforradalom jelentett a Big Tech cégek és a pénzügyi világ számára, csak lassan szivárog át a nem intenzíven finanszírozott területekre, mint például a társadalomtudományok. Jelenleg viszonylag alacsony ezen a területen a Big Data-t is felhasználó kutatások száma, a módszertana szintén kidolgozatlan. Ebből kifolyólag nagy hangsúlyt kap az egyéni kreativitás és intuíció. Azon kutatóknak, akik ebbe a típusú elemzésbe

belevágnak, több területen is átfogó ismeretekkel kell rendelkezniük, valamint hozzáférést kell szerezniük az ingyenesen és/vagy fizetés ellenében elérhető információkhoz.

A módszerben rejlő lehetőségek szinte korlátlanok, általa eddig nem kutatott (és nem is dokumentált) dimenziókhoz férhetünk hozzá. Az adatelemzéssel az eredmények gyorsan tesztelhetővé, pontosabbá és megbízhatóbbá tehetőek. A hagyományosnak tekintett módszerek, a kisebb/nagyobb mintás felmérések és a minőségi elemzések továbbra is fontosak, az általánosan elvárt kutatói etika továbbra is érvényes marad.

BIBLIOGRÁFIA

Albert-Lőrincz Cs., Belényi Emese, (2019), Felkészülés és képzéshasznosulás a segítő szakmákban. Összehasonlító tanulmány, Párbeszéd 6(2):19 o. https://ojs.lib.unideb.hu/parbeszed/article/view/3257/3221 (hozzáférés: 2023.01.22)

Albert-Lőrincz Cs. (2022). A big data elemzések perspektívái a társadalomtudományi kutatásokban. Humán Innovációs Szemle. XIII(2): 25-32. http://humanexchange.hu/site/uploads/file/HISZ_2022_2_.pdf (hozzáférés: 2023.08.29)

Al-Khouri A. M. (2012): Data Ownership: Who Owns ‚My Data'? *International Journal of Management & Information Technology* 2(1). pp. 1-8. DOI: 10.24297/ijmit.v2i1.1406

Boncz Imre (2015). Kutatásmódszertani alapismeretek, Pécsi Tudományegyetem Egészségtudományi Kar, Pécs.

Barbour, R.S., Kitzinger, J. (1999): *Developing focus group research. Politics, theory and practice*. Sage Publications. https//doi.org/10.4135/9781849208857

Bujang, M. A., Baharum, N. A. (2017). Guidelines of the minimum sample size requirements for Cohen's Kappa. Epidemiology Biostatistics and Public Health. 14(2): e12267-1 - e12267-10. doi:10.2427/12267

Chai W., Labbe M. & Stedman C. (s.a.): Big data analytics. *Techtarget.* https://www.techtarget.com/searchbusinessanalytics/definition/big-data-analytics (hozzáférés: 2023.02.02)

Central Michigan University (s.a.): Projects Using Cars Data. http://calcnet.mth.cmich.edu/org/spss/prj_carsdata.htm (hozzáférés: 2023.02.02).

Choi H. & Hal V. (2009): *Predicting the Present with Google Trends.* Google Inc. https://static.googleusercontent.com/media/www.google.com/en//googleblogs/pdfs/google_predicting_the_present.pdf (hozzáférés: 2023.08.29)

Compagnucci M. C. (2020): *Big Data, Databases and „Ownership" Rights in the Cloud*. Springer Singapore, DOI:10.1007/978-981-15-0349-8

Desrosières, A. (2016): *The quantification of the social sciences: an historical comparison.* Springer Cham. DOI:10.1007/978-3-319-44000-2_15

Európai Parlament és Tanács. (2016): Az Európai Parlament és a Tanács (EU) 2016/679 számú Rendelete a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről (általános adatvédelmi rendelet). *Az Európai Unió Hivatalos Közlönye* 59, pp. 1-88. https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=CELEX:32016R0679&from=EN (hozzáférés: 2023.08.29)

Fadler, M. & Legner, C. (2020): *Who owns data in the enterprise? Rethinking data ownership in times of Big Data and analytics.* Twenty-Eigth European Conference on Information Systems (ECIS2020). https://serval.unil.ch/resource/serval:BIB_E6BE23FE6102.P001/REF (hozzáférés: 2023.08.29)

Frost, J. (é.n.). Nonparametric Tests vs. Parametric Tests, Statisticsbyjim.com, https://statisticsbyjim.com/hypothesis-testing/nonparametric-parametric-tests/ (hozzáférés: 2023.01.16)

Hal V. (2011): *Predicting the Present*. Google Inc., https://www.thinkwithgoogle.com/marketing-strategies/search/predicting-the-present (hozzáférés: 2023.08.29)

Haldun Akoglu (2018). User's guide to correlation coefficients. Turkish Journal of Emergency Medicine 18(3): 91-93. https://doi.org/10.1016/j.tjem.2018.08.001

Hesse B. W., Moser R.P. & Riley W. T. (2015): From Big Data to Knowledge in the Social Sciences. *The ANNALS of the American Academy of Political and Social Science* 659(1). pp. 16-32. DOI:10.1177/0002716215570007

Houser, K. A. & Voss, W. G. (2018): GDPR: The End of Google and Facebook or a New Paradigm in Data Privacy? *Richmond Journal of Law & Technology*, 25(1). p. 109.

Kutrovátz Kitti (2016). Kvalitatív interjúzás kamaszokkal – Módszertani és etikai kihívások a gyerekek kutatásában, Prosperitas, 3(2): 88–110.

Kvale, S. (2005): *Az interjú. Bevezetés a kvalitatív kutatás interjútechnikáiba*. Ford. Kovács B., Sivadó Á., Jószöveg Műhely Kiadó, Budapest

LaValley, M. P. (2008). Logistic Regression. Circulation. 117:2395–2399. doi.org/10.1161/CIRCULATIONAHA.106.682658

Lohr S. (2012): The age of Big Data. *The New York Times*. https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html (hozzáférés: 2023.08.29)

Malatyinszki Sz. (2020): Experiencing digitaleducation. EasyChair 3674, 1-7. https://easychair-www.easychair.org/publications/preprint/GtvW (hozzáférés: 2023.08.29)

Mezei E., Veres V. (2001): Társadalomstatisztika. Egyetemi jegyzet. Kolozsvári Egyetemi Kiadó, Kolozsvár.

McDermott, Y. (2017): Conceptualising the right to data protection in an era of Big Data. *Big Data & Society*, 4(1). pp. 1-7. https://doi.org/10.1177/2053951716686994

McNeely, C. L. & Jong-on H. (2014): The Big (Data) Bang: Policy, Prospects, and Challenges. *Review of Policy Research*, 31(4). pp. 304-310.

Naqshbandi, G. (2021): "Big Data: The 3 V's of Data." *Wevolver*, 2021:9, www.wevolver.com/article/big-data-the-3-v-s-of-data

Nemzetközi Szociológiai Társaság (ISA), 2001, Etikai Kódex. https://www.isa-sociology.org/en/about-isa/code-of-ethics (hozzáférés: 2023.03.15-én)

Nielsen M. (2013): Who Owns Big Data? In *Ch@nge: 19 Key Essays on How the Internet Is Changing Our Lives.* Madrid: BBVA. https://www.bbvaopenmind.com/en/articles/who-owns-big-data

Ovadia, S. (2013) The Role of Big Data in the Social Sciences, Behavioral & Social Sciences Librarian, 32:2, 130-134, DOI: 10.1080/01639269.2013.787274

Pallant, J. (2007). SPSS Survival Manual. Open University Press. Berkshire

Partiumi Keresztény Egyetem, é.n., Magyar online adatbázisok, https://htt.partium.ro/hu/hasznos-informaciok/magyar-online-adatbazisok (hozzáférés: 2020.03.23)

PSPP Users' Guide (2020). Free Software Foundation, Inc. https://www.gnu.org/software/pspp/manual/pspp.pdf (hozzáférés: 2023.01.22)

Runkel, P. J. (1962): Quantification in the social sciences. *The Mathematics Teacher*, *55*(1). pp. 20–33. http://www.jstor.org/stable/27956514 (hozzáférés: 2023.08.29)

Sim, J., Wright, Ch. C (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, 85(3): 257-268, https://doi.org/10.1093/ptj/85.3.257

SPSSABC.HU. (é.n). A normális eloszlás jellemzői és vizsgálata. https://spssabc.hu/a-normalis-eloszlas-jellemzoi-es-vizsgalata/ (hozzáférés: 2022.01.12-én)

Statista.com. (2022): Worldwide desktop market share of leading search engines from January 2010 to July 2022. https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/#:~:text=Google%20in%20the%20global%20market&text=Google%20has%20dominated%20the%20search,share%20as%20of%20June%202021 (hozzáférés: 2023.08.29)

Stephenz-Davidowitz S. (2019): *Mindenki hazudik.* Athenaeum Kiadó, Budapest

Szűts Z. & Yoo J. (2016): Big Data, az információs társadalom új paradigmája. *Információs Társadalom*. 16(1). pp. 8-28. DOI: 10.22503/inftars.XVI.2016.1.1

Tabachnick, G. B., Fidell, L. S. (2013). Using multivariate statistics (6th ed.). Pearson Education. Upper Saddle River.

Tesfay, W. B., Hofmann P., Nakamura T. et. al. (2018): PrivacyGuide: Towards an Implementation of the EU GDPR on Internet Privacy Policy Evaluation. In: *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics (IWSPA ,18).* Association for Computing Machinery, New York. pp. 15–21. DOI: 10.1145/3180445.3180447

van Ooijen, I.; Vrabec, H. U. (2019): Does the GDPR Enhance Consumers' Control over Personal Data? An Analysis from a Behavioural Perspective. *Journal of Consumer Policy*, 2019(42). pp. 91–107. DOI:10.1007/s10603-018-9399-7

van Smeden, M., de Groot, J.A., Moons, K.G. *et al.* (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Med Res Methodol.* 16:163. 12 o. https://doi.org/10.1186/s12874-016-0267-3

Wiltshire D. & Alvanides S. (2022): Ensuring the ethical use of big data: lessons from secure data access. *Heliyon,* 8(2). pp. 1-6. DOI:10.1016/j.heliyon.2022.e08981.

Zarsky, Tal Z. (2017): Incompatible: The GDPR in the Age of Big Data. *The Seton Hall Law Review,* 47(2). pp. 995-1020.



ISBN: 978-606-37-2017-8